Report No. 244

VARIABLE SUCCESSIVE OVER-RELAXATION

by

Leland K. McDowell

September 18, 1967

DEPARTMENT OF COMPUTER SCIENCE · UNIVERSITY OF ILLINOIS · URBANA, ILLINOIS

Report No. 244

VARIABLE SUCCESSIVE OVER-RELAXATION*

by

Leland K. McDowell

September 18, 1967

Department of Computer Science
University of Illinois
Urbana, Illinois 61801

[*] This work was submitted in partial fulfillment of the Doctor of Philosophy in Mathematics and was supported in part by the National Science Foundation under Grant No. NSF-GP-4636.

ACKNOWLEDGMENT

The author ackncwledges with pleasure the assistance and encouragement of Professor Donald B. Gillies, in whose researches the problem treated here originated and whose many helpful comments and suggestions contributed greatly to the present work.

The author would also like to express his gratitude to the Department of Computer Science of the University of Illinois for providing him with financial support during the preparation of this work and for making available to him the facilities of the ILLIAC II computer system. This work was supported in part by the National Science Foundation under Grant No. GP-4636.

Many thanks are also due to Miss Bonnie J. Malcor for patient and careful typing of the manuscript.

TABLE OF CONTENTS

VARIABLE SUCCESSIVE OVER-RELAXATION

by

Leland Kitchin McDowell
Department of Mathematics
University of Illinois, 1967

This thesis investigates the solution of linear systems by extrapolated Gauss-Seidel iteration using a multiplicity of extrapolation parameters. This is a generalization of the method of successive over-relaxation, which uses a single scalar extrapolation parameter. The linear systems considered are those which arise in the numerical solution of boundary value problems for self-adjoint, elliptic partial differential equations.

In Chapter 2 it is shown that the use of two appropriately chosen scalar extrapolation factors yields an iteration having a higher rate of convergence than SOR, and formulas are derived for choosing the factors optimally. Also, it is shown that by the use of extrapolation matrices, an iteration can be constructed whose matrix is nilpotent, i.e., whose rate of convergence is infinite.

Chapter 3 considers a more limited class of linear systems for which a certain sort of spectral decomposition is possible. For the solution of such systems the SEI and VSEI methods are introduced and shown to be equivalent to several simultaneous extrapolated Gauss-Seidel iterations on certain subspaces, each with a different extrapolation factor or set of factors. Theoretical and experimental results are presented which show that SEI, which requires less work per iteration than SOR, has the same asymptotic rate of convergence and, for most starting vectors, an improved actual rate of convergence. Experimental evidence is presented showing that certain versions of VSEI have a higher asymptotic rate of convergence for the problems considered than SOR.

1. INTRODUCTION

1.1 Relaxation Methods

The use of relaxation for the solution of linear systems is documented as early as 1823, when Gauss [3][*] wrote favorably of his experience with such a procedure, which he called indirect elimination. The Jacobi Iteration for the solution of a linear system

$$A\vec{Z} = \vec{K} \tag{1.1.1}$$

where A is symmetric was discussed in 1846 by Jacobi [7], who used plane rotations to increase the diagonal dominance of A and hence to improve convergence.

In 1862, Seidel [10] considered the method now known as Gauss-Seidel iteration for the case in which A is m x n with $m \geq n$. Seidel proved that his iteration applied to

$$A^T A \vec{Z} = A^T \vec{K} \tag{1.1.2}$$

converges to a point which best satisfies (1.1.1) in the sense of least squares and hence to the unique solution if A is square and non-singular.

Seidel noted that if there exists a subsystem of k unknowns and k equations in which the unknowns within the subsystem are coupled to only a few outside unknowns, then it is profitable to relax repeatedly the residuals of the subsystem until they become small while treating the outside unknowns as constants. Thus he anticipated modern block relaxation methods, in which a number of residuals are relaxed simultaneously.

[*] Numbers in square brackets refer to items in the List of References.

Both the Jacobi and the Gauss-Seidel iterations are known to be convergent if, for example, the coefficient matrix is either strictly or irreducibly diagonally dominant [5].

Experience with hand computation showed that it is often profitable to over-relax, and intuition indicates that the greatest advantage lies in relaxing the largest residual at each stage. If this is done, then the judgment of the relaxer intervenes to alter the algorithm from one iteration to the next, making the iteration non-cyclic.

The use of electronic digital computers permits the application of relaxation methods to linear systems of very large order, but such machines are best suited to cyclic iterative methods, in which the algorithm once defined is repeated without alteration at each iteration. For linear systems arising in the discretization of boundary value problems for self-adjoint, elliptic partial differential equations, Frankel [4] and Young [12] in 1950 described such a cyclic procedure for extrapolated Gauss-Seidel iteration, often called successive over-relaxation (SOR). Young showed that for certain orderings of the unknowns, which he termed consistent orderings, a functional relationship exists between the extrapolation factor ω and the rate of convergence of SOR. Using this relationship, Young derived formulas for the optimum value of ω and for the rate of convergence of the resulting iteration, which he showed to be substantially faster than Gauss-Seidel iteration, especially when the latter is slow.

Young's results, which applied to point iterations, were extended to block iterations in 1956 by Arms, Gates, and Zondek [1].

Few results are known concerning the use of extrapolation in case no consistent ordering of the unknowns exists, but Ostrowski [9] proved in 1954 that if A is Hermitian, then SOR converges if and only if A is positive definite and $0 < \omega < 2$.

In 1959 Golub [6] introduced the Chebyshev semi-iterative method and the modified (or cyclic) Chebyshev semi-iterative method. The latter is a variation of SOR using a particular consistent ordering of the unknowns and a sequence of extrapolation factors, a different pair of factors being used for each iteration. The cyclic Chebyshev semi-iterative method provides an improved average rate of convergence over SOR, although the asymptotic rates of convergence of the two iterations are the same.

1.2 Variable Extrapolation Factors

Whereas SOR uses the same scalar extrapolation factor for each unknown of the linear system being solved, this thesis investigates extrapolated Gauss-Seidel iteration using extrapolation factors which vary from one unknown to another or from one block of unknowns to another. The use of matrices as extrapolation parameters is also investigated. We call such iterations variable successive over-relaxation (VSOR)

The linear systems considered are those for which the use of a single extrapolation factor is known to be profitable, e.g., linear systems arising from the discretization of boundary value problems for self-adjoint elliptic partial differential equations

Chapter 2 begins by introducing *2-factor VSOR*, an extrapolated Gauss-Seidel iteration using two scalar extrapolation factors. Formulas for the optimum pair of factors are obtained, and it is shown that the rate of convergence of 2-factor VSOR is greater than that of SOR.

In Section 2 of Chapter 2 the concept of *extrapolation matrices* is introduced, and it is shown that by their use an extrapolated Gauss-Seidel iteration can be constructed whose iteration matrix is nilpotent. Hence, in the absence of rounding errors, the exact solution is obtained after a finite number of iterations. Moreover, it is shown that if the non-null blocks of the associated block Jacobi matrix are all square of order r and have a common basis of eigenvectors, then VSOR with extrapolation matrices is equivalent to r simultaneous extrapolated Gauss-Seidel iterations each using a sequence of scalar extrapolation factors. Thus, a decomposition theorem for VSOR is obtained.

In Chapter 3 we consider the linear system resulting from the discretization of the Dirichlet problem for Poisson's equation on a rectangle. The resulting rq x rq linear system is the model problem for whose solution the *sequential extrapolated implicit method (SEI)* is introduced. This method accelerates Gauss-Seidel iteration by means of a shift of origin, which requires less work per iteration than extrapolation by a scalar. It is shown, however, that SEI is actually equivalent to the use of certain extrapolation matrices, and the decomposition theorem of Chapter 2 is then applied to show that SEI is equivalent to performing SOR simultaneously in each of r subspaces, a different scalar extrapolation factor being used in each subspace. It is then shown that the asymptotic rates of convergence of SEI and SOR are identical when the optimum acceleration parameter is used for each, but that the actual rate of convergence of SEI is always at least as high as that of SOR and for certain starting vectors is substantially higher.

In Section 3.3 *cyclic Chebyshev SEI* is introduced The relationship of this iteration to SEI is analogous to the relationship of the cyclic Chebyshev semi-iterative method to SOR.

Section 3.4 presents numerical results comparing the actual rates of convergence using various starting vectors of SOR, SEI, the Chebyshev semi-iterative method, and cyclic Chebyshev SEI.

Finally, the *variable sequential extrapolated implicit method (VSEI)* for the solution of the model problem is introduced. Several versions of this iteration are investigated, each of which uses the criteria of Chapter 2 for the construction of nilpotent iteration matrices together with the decomposition theorem to make the VSEI iteration matrix nilpotent on certain q-dimensional subspaces of rq-space. Experimental evidence is presented which shows that VSEI is asymptotically faster than SOR for the model problem.

VARIABLE SUCCESSIVE OVER-RELAXATION FOR LINEAR SYSTEMS WHOSE COEFFICIENT MATRICES ARE CONSISTENTLY ORDERED S-MATRICES

2.1 Basic Concepts

Throughout this thesis we will be concerned with the iterative solution of the linear system

$$A\vec{Z} = \vec{K} \tag{2.1.1}$$

where A is a matrix usually of large order and sparse; i.e., having only few non-zero entries. A detailed development of the ideas outlined in this section may be found in [11, Chap. 1,3] or [13, Chap. 1].

Our results will apply in particular in case A is an S-matrix; i e., A satisfies

(i) A is real and symmetric

(ii) $a_{ij} \leq 0$ for $i \neq j$

(iii) A is irreducible; i.e., there exists no permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

where the diagonal submatrices are square

(iv) A is diagonally dominant; i.e.,

$$a_{ii} \geq \sum_{j \neq i} |a_{ij}| , \qquad \forall_i$$

(v) The inequality in (iv) is strict for at least one i.

S-matrices arise, for example, in the discretization of self-adjoint, elliptic partial differential equations.

A stationary, linear iteration for the solution of (2.1.1) is an <u>affine</u> transformation

$$\vec{Z}^{m+1} = H\vec{Z}^{m} + \vec{F} \tag{2.1.2}$$

with the property that

$$\vec{Z} = H\vec{Z} + \vec{F} \tag{2.1.3}$$

The error in the m-th iterate is $\vec{E}^{m} = \vec{Z} - \vec{Z}^{m}$, and subtraction of (2.1.3) from (2.1.2) shows that

$$\vec{E}^{m+1} = H\vec{E}^{m} = H^{m+1}\vec{E}^{o} \tag{2.1.4}$$

Hence, the convergence properties of the iteration (2.1.2) depend only upon the matrix H. In particular, $\lim_{m\to\infty} \vec{E}^{m} = \vec{O}$ for arbitrary $\vec{E}^{o}$ if and only if $\lim_{m\to\infty} H^{m}$ is the null matrix. This is the case if and only if all eigenvalues of H have modulus less than one. The <u>spectral radius</u> of H, denoted by $\rho(H)$, is the maximum of the moduli of the eigenvalues of H. and the (<u>asymptotic</u>) <u>rate of convergence</u> of the iteration (2.1.2) is $R = -\ln \rho$.

A partitioning of A in which the i-th diagonal submatrix A_{ii} is $r_i \times r_i$ induces a partitioning $\vec{Z}^{T} = [\vec{Z}_1^{T} \mid \vec{Z}_2^{T} \mid \cdots \mid \vec{Z}_q^{T}]$ of the unknown vector $\vec{Z}$ such that the block $\vec{Z}_i$ contains r_i of the unknowns. The constant vector $\vec{K}$ is similarly partitioned.

If A is an S-matrix, then A is positive definite. Each A_{ii} is therefore also positive definite and hence non-singular. The <u>block Jacobi</u> iteration corresponding to the particular partitioning of A is

$$\vec{Z}_i^{m+1} = -\sum_{j\neq i} A_{ii}^{-1} A_{ij} \vec{Z}_j^{m} + A_{ii}^{-1} \vec{K}_i, \quad 1 \leq i \leq q, \tag{2.1.4}$$

where the partitioning is assumed to be such that A_{ij} is $r_i \times r_j$. Throughout this thesis M will denote the matrix for the iteration (2.1.4); i.e., _M denotes the block Jacobi matrix derived from A_.

The block Gauss-Seidel iteration differs from the block Jacobi in that as soon as a block $\vec{Z}_i^{m+1}$ of the next vector iterate is calculated, it and not $\vec{Z}_i^m$ is used in the computation of later blocks $\vec{Z}_{i+1}^{m+1}$, $\vec{Z}_{i+2}^{m+1}$, etc. Thus,

$$\vec{Z}_i^{m+1} = - \sum_{j<i} A_{ii}^{-1} A_{ij} \vec{Z}_j^{m+1} - \sum_{j>i} A_{ii}^{-1} A_{ij} \vec{Z}_j^{m} + A_{ii}^{-1} \vec{K}_i , \quad 1 \leq i \leq q. \tag{2.1.5}$$

If P_σ is a permutation matrix, then the coordinate transformation

$$P_\sigma A P_\sigma^T \vec{Z} = P_\sigma \vec{K} \tag{2.1.6}$$

corresponds to a reordering of the unknowns. The ordering σ is consistent if $A_\sigma \equiv P_\sigma A P_\sigma^T$ is block tri-diagonal (i.e., A_{ij} is null if $|i-j|>1$) with square diagonal blocks, and if so, then A_σ is said to be _consistently ordered_. The block Jacobi matrix M_σ derived from A_σ is said to be consistently ordered if A_σ is. Throughout this thesis A and M will be assumed consistently ordered already and the subscript σ will be omitted.

If M is consistently ordered, it is of the form

$$\begin{bmatrix} 0 & A_{11}^{-1}\,A_{12} & & & \bigcirc \\ A_{22}^{-1}\,A_{21} & 0 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 0 & A_{q-1q-1}\;A_{q-1q} \\ \bigcirc & & & A_{qq}^{-1}\,A_{q,q-1} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & B_{12} & & & \bigcirc \\ B_{21} & 0 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 0 & B_{q-1q} \\ \bigcirc & & & B_{qq-1} & 0 \end{bmatrix}$$

so $B_{ij} = 0$ for $|i-j| \neq 1$. Hence $M = L + U$, where L is strictly lower triangular and U is strictly upper triangular. Moveover, if M is consistently ordered and μ is an eigenvalue, then $-\mu$ is an eigenvalue of the same multiplicity.

If

$$D = \begin{bmatrix} A_{11} & & \bigcirc \\ & \ddots & \\ \bigcirc & & A_{qq} \end{bmatrix}$$

then $M = I - D^{-1} A$. Hence M is similar to $I - D^{-1/2} AD^{-1/2}$, and it follows that this is symmetric because the square root of a positive definite symmetric matrix is symmetric. Hence all of the eigenvalues of M are real.

2.2 Successive Over-Relaxation (SOR)

Define the intermediate value

$$\vec{Z}_i^{m+\frac{1}{2}} = B_{i,i-1} \vec{Z}_{i-1}^{m+1} + B_{i,i+1} \vec{Z}_{i+1}^{m} + A_{ii}^{-1} \vec{K}_i \tag{2.2.1}$$

and extrapolate by a fixed scalar ω:

$$\vec{Z}_i^{m+1} = \omega[\vec{Z}_i^{m+\frac{1}{2}} - \vec{Z}_i^{m}] + \vec{Z}_i^{m}, \quad 1 \leq i \leq q . \tag{2.2.2}$$

Then

$$\begin{aligned} \vec{Z}_i^{m+1} &= \omega[B_{i,i-1} \vec{Z}_{i-1}^{m+1} + B_{i,i+1} \vec{Z}_{i+1}^{m}] + (1-\omega) \vec{Z}_i^{m} \\ &\quad + \omega A_{ii}^{-1} \vec{K}_i, \quad 1 \leq i \leq q. \end{aligned} \tag{2.2.3}$$

The iteration (2.2.3) is the <u>successive over-relaxation method with extrapolation factor ω</u>. The matrix for the iteration (2.2.3) is

$$\mathcal{L}_\omega = [I - \omega L]^{-1} [\omega U + (1 - \omega)I] .$$

Let $0 < \mu_1 \leq \mu_2 \leq \cdots \leq \mu_\ell = \bar{\mu}$ be the positive eigenvalues of M, and let p be the multiplicity of zero as an eigenvalue. For $1 \leq i \leq \ell$, let

$$\varphi_i(\lambda) = \lambda^2 - [2(1-\omega) + \omega^2 \mu_i^2] \lambda + (1-\omega)^2 .$$

Then the characteristic polynomial of $\mathcal{L}_\omega$ is

$$\Phi(\lambda) = [(1 - \omega) - \lambda]^p \prod_{i=1}^{\ell} \varphi_i(\lambda) .$$

The maximum of the moduli of the roots of $\varphi_\ell(\lambda)$ is minimized as a function of ω when the roots are real and equal; that is, when the discriminant of $\varphi_\ell(\lambda)$ is zero. This is the case when

$$\omega = \omega_b = \frac{2}{1 + \sqrt{1 - \bar{\mu}^2}} .$$

For $\omega = \omega_b$, the modulus of the roots is

$$\rho_{SOR} = \frac{1 - \sqrt{1 - \bar{\mu}^2}}{1 + \sqrt{1 - \bar{\mu}^2}} .$$

If the discriminant of $\varphi_\ell(\lambda)$ is negative or zero for a particular value of ω, then the discriminant of $\varphi_i(\lambda)$, $1 \leq i < \ell$, is also negative or zero. It follows that for $\omega = \omega_b$ the roots of $\Phi(\lambda)$ all lie on a circle of radius ρ_{SOR}, and hence

$$\min_\omega \rho(\mathcal{L}_\omega) = \rho(\mathcal{L}_{\omega_b}) = \rho_{SOR} .$$

This completes the summary of known results.

2.3 VSOR With Two Extrapolation Factors

In this section we consider a linear, stationary iteration using two distinct extrapolation factors. It will be shown that for certain consistently ordered S-matrices an optimally chosen pair of factors used in alternation on consecutive blocks yield an iteration having a higher rate of convergence than SOR.

For any two non-zero scalars ω_1 and ω_2 consider the iteration

$$\vec{Z}_i^{m+1} = \omega_i [B_{i,i-1} \vec{Z}_{i-1}^{m+1} + B_{i,i+1} \vec{Z}_{i+1}^{m}] + (1 - \omega_i) \vec{Z}_i^{m} + \omega_i A_{ii}^{-1} \vec{K}_i, \; 1 \leq i \leq q, \tag{2.3.1}$$

where

$$\omega_i = \begin{Bmatrix} \omega_1 & \text{if } i \text{ is odd} \\ \omega_2 & \text{if } i \text{ is even} \end{Bmatrix}$$

Let $\mathfrak{L}_{\omega_1,\omega_2}$ denote the iteration matrix for (2.3.1), and suppose that $\vec{X}$ partitioned like $\vec{Z}$ is an eigenvector of $\mathfrak{L}_{\omega_1,\omega_2}$ belonging to the eigenvalue λ. It follows from (2.3.1) that

$$\lambda \vec{X}_i = \omega_i [\lambda B_{i,i-1} \vec{X}_{i-1} + B_{i,i+1} \vec{X}_{i+1}] + (1 - \omega_i) \vec{X}_i, \; 1 \leq i \leq q,$$

or, after dividing by ω_i,

$$0 = [\lambda B_{i,i-1} \vec{X}_{i-1} + B_{i,i+1} \vec{X}_{i+1}] + [\frac{(1 - \omega_i) - \lambda}{\omega_i}] \vec{X}_i \; .$$

This linear transformation on $\vec{X}$ we write as $\check{G} \vec{X} = \vec{0}$. If

$$\check{\gamma}_i = \frac{(1 - \omega_i) - \lambda}{\omega_i} \; ,$$

then

$$\check{G} = \begin{bmatrix} \check{\gamma}_1 I_1 & B_{12} & & \bigcirc \\ \lambda B_{21} & \check{\gamma}_2 I_2 & \ddots & \\ & \ddots & \ddots & B_{q-1,q} \\ \bigcirc & & \lambda B_{q,q-1} & \check{\gamma}_q I_q \end{bmatrix}$$

Here I_i is the identity matrix of the same order as A_{ii}.

The equation $\check{G}\,\vec{X} = \vec{0}$ has a non-trivial solution if and only if det $(\check{G}) = 0$, and since det $(\check{G})$ is a polynomial of degree n in λ, its roots are the eigenvalues of $\mathfrak{L}_{\omega_1,\omega_2}$. Assuming $\lambda \neq 0$ (an assumption which will later be justified), we premultipy $\check{G}$ by

$$\Lambda_\ell = \begin{bmatrix} \lambda^{-\frac{1}{2}} I_1 & & \bigcirc \\ & \lambda^{-1} I_2 & \\ & & \ddots & \\ \bigcirc & & & \lambda^{-\frac{q}{2}} I_q \end{bmatrix}$$

and postmultiply by

$$\Lambda_r = \begin{bmatrix} I_1 & & \bigcirc \\ & \lambda^{\frac{1}{2}} I_2 & \\ & & \ddots & \\ \bigcirc & & & \lambda^{\frac{q-1}{2}} I_q \end{bmatrix}$$

This transformation multiplies each subdiagonal block by λ^{-1}, each diagonal block by $\lambda^{-\frac{1}{2}}$, and each superdiagonal block by 1. Hence, letting $\gamma_i = \lambda^{-\frac{1}{2}} \check{\gamma}_i$, we have

$$G \equiv \Lambda_\ell \check{G} \Lambda_r = \begin{bmatrix} \gamma_1 I_1 & B_{12} & & \bigcirc \\ B_{21} & \gamma_2 I_2 & \ddots & \\ & \ddots & \ddots & B_{q-1,q} \\ \bigcirc & & B_{q,q-1} & \gamma_q I_q \end{bmatrix}$$

But det (Λ_ℓ) det $(\Lambda_r) = \lambda^{-\frac{n}{2}}$, and so the following lemma has been proved.

Lemma 2.3.1: $\lambda \neq 0$ is an eigenvalue of $\mathcal{L}_{\omega_1,\omega_2}$ if and only if it is a zero of det (G).

Since A_{ii} is of order r_i, the vector $\vec{Z}_i$ contains r_i components, and it follows that ω_1 is used with $r_o = \sum_{i \text{ odd}} r_i$ components and ω_2 is used with $r_e = \sum_{i \text{ even}} r_i$ components.

Lemma 2.3.2: Let the matrix G' be obtained by replacing each diagonal element γ_i of G by $(\gamma_1 \gamma_2)^{1/2}$. Then

$$\det(G) = \gamma_1^{\frac{r_o - r_e}{2}} \gamma_2^{\frac{r_e - r_o}{2}} \det(G') .$$

Proof:

Let

$$\Gamma_r = \begin{bmatrix} \gamma_1^{\frac{1}{2}} I_1 & & & \bigcirc \\ & \gamma_2^{\frac{1}{2}} I_2 & & \\ & & \gamma_1^{\frac{1}{2}} I_3 & \\ \bigcirc & & & \ddots \end{bmatrix}$$

and

$$\Gamma_\ell = \begin{bmatrix} \gamma_2^{-\frac{1}{2}} I_1 & & & \bigcirc \\ & \gamma_1^{-\frac{1}{2}} I_2 & & \\ & & \gamma_2^{-\frac{1}{2}} I_3 & \\ \bigcirc & & & \ddots \end{bmatrix}$$

Then $G = \Gamma_\ell G' \Gamma_r$, and so $\det(G) = \det(\Gamma_\ell) \det(G') \det(\Gamma_r)$. But

$\det(\Gamma_\ell) \det(\Gamma_r) = \gamma_1^{\frac{r_o - r_e}{2}} \gamma_2^{\frac{r_e - r_o}{2}}$, which proves the lemma.

Lemma 2.3.3: Let $\mu_1, \mu_2, \ldots, \mu_\ell$ be the positive eigenvalues of M, and let p denote the multiplicity of zero as an eigenvalue. For $1 \leq i \leq \ell$, define

$$\varphi_{\mu_i}(\lambda) = \lambda^2 - [(1 - \omega_1) + (1 - \omega_2) + \omega_1\omega_2\mu_i^2]\,\lambda + (1 - \omega_1)(1 - \omega_2).$$

Then the zeroes of det (G') are precisely the zeroes of

$$[(1 - \omega_1) - \lambda]^{\frac{p}{2}}\,[(1 - \omega_2) - \lambda]^{\frac{p}{2}}\,\prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda)\ .$$

Proof:

$$G' = \begin{bmatrix} \gamma_1^{\frac{1}{2}}\,\gamma_2^{\frac{1}{2}}\,I_1 & B_{12} & & \bigcirc \\ B_{21} & \gamma_1^{\frac{1}{2}}\,\gamma_2^{\frac{1}{2}}\,I_2 & \ddots & \\ & \ddots & \ddots & B_{q-1,q} \\ \bigcirc & & B_{q,q-1} & \gamma_1^{\frac{1}{2}}\,\gamma_2^{\frac{1}{2}}\,I_q \end{bmatrix}$$

is, except for the diagonal blocks $\gamma_1^{1/2}\,\gamma_2^{1/2}\,I_j$, identical to M. Hence $\det(G') = 0$ if and only if $\gamma_1^{1/2}\,\gamma_2^{1/2} = \mu_i$, where μ_i is an eigenvalue of M. Moreover, the multiplicity of $(\gamma_1^{1/2}\,\gamma_2^{1/2} - \mu_i)$ as a factor of det (G') is equal to the multiplicity of μ_i as an eigenvalue of M. If $\mu_i \neq 0$, then $(\gamma_1^{1/2}\,\gamma_2^{1/2} + \mu_i)$ is also a factor of det (G') and therefore so is

$(\gamma_1 \gamma_2 - \mu_i^2)$. But $\gamma_1 \gamma_2 - \mu_i^2 = \dfrac{[(1-\omega_1) - \lambda][(1-\omega_2) - \lambda]}{\omega_1 \omega_2 \lambda} - \mu_i^2$ from which it follows that corresponding to the 2ℓ non-zero eigenvalues of M are 2ℓ zeroes of det(G), namely the 2ℓ roots of $\prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda) = 0$.

The remaining zeroes of det(G) are the zeroes of

$$(\gamma_1^{\frac{1}{2}} \gamma_2^{\frac{1}{2}} - 0)^p = \left\{\frac{[(1-\omega_1) - \lambda][(1-\omega_2) - \lambda]}{\omega_1 \omega_2 \lambda}\right\}^{\frac{p}{2}}$$

from which the conclusion of the lemma follows.

<u>Lemma 2.3.4</u>: $p_1 = \dfrac{p + (r_o - r_e)}{2}$ and $p_2 = \dfrac{p + (r_e - r_o)}{2}$ are both non-negative integers.

<u>Proof</u>: $r_o + r_e = n = p + 2\ell$. Hence, $(r_o + r_e)$ has the same parity as p, and so, therefore, does $r_o - r_e$. Then $p + (r_o - r_e)$ and $p - (r_o - r_e)$ are both even, and so p_1 and p_2 are integers.

By lemmas 2.3.2 and 2.3.3, the zeroes of det(G) are precisely the zeroes of $[(1 - \omega_1) - \lambda]^{p_1}[(1 - \omega_2) - \lambda]^{p_2} \prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda)$. There are exactly n such zeroes because each one is an eigenvalue of $\mathcal{L}_{\omega_1, \omega_2}$ by lemma 2.3.1. Hence $p_1 \leq p$ and $p_2 \leq p$. But $p_1 + p_2 = p$, so both p_1 and p_2 are non-negative.

<u>Theorem 2.3.1</u> The characteristic polynomial of $\mathcal{L}_{\omega_1, \omega_2}$ is

$$\Phi(\lambda) = [(1 - \omega_1) - \lambda]^{p_1}[(1 - \omega_2) - \lambda]^{p_2} \prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda)$$

<u>Proof</u>: By lemmas 2.3.1, 2.3.2, and 2.3.3, the zeros of $\Phi(\lambda)$ are precisely the eigenvalues of $\mathcal{L}_{\omega_1, \omega_2}$. But by lemma 2.3.4 $\Phi(\lambda)$ is a polynomial, which proves the theorem.

In practical problems the eigenvalues of M are usually not known, but an interval can be determined in which the eigenvalues are known to lie. We therefore consider choosing ω_1 and ω_2 to minimize the maximum of the moduli of the roots of

$$\varphi_\mu(\lambda) = \lambda^2 - [(1 - \omega_1) + (1 - \omega_2) + \omega_1\omega_2\mu^2]\lambda + (1 - \omega_1)(1 - \omega_2)$$

where it is known only that $0 \leq \underline{\mu} \leq \mu \leq \bar{\mu} < 1$.

The discriminant of $\varphi_\mu(\lambda)$ is

$$D(\mu) = \omega_1\omega_2\mu^2[\omega_1\omega_2\mu^2 + 2(1 - \omega_1) + 2(1 - \omega_2)] + (\omega_2 - \omega_1)^2.$$

Lemma 2.3.5: If $\omega_1 > 1$, $\omega_2 > 1$, $D(\underline{\mu}) \leq 0$, and $D(\bar{\mu}) \leq 0$, then $D(\mu) \leq 0$ for $\underline{\mu} \leq \mu \leq \bar{\mu}$.

Proof: This is evident from the graph of $D(\mu)$.

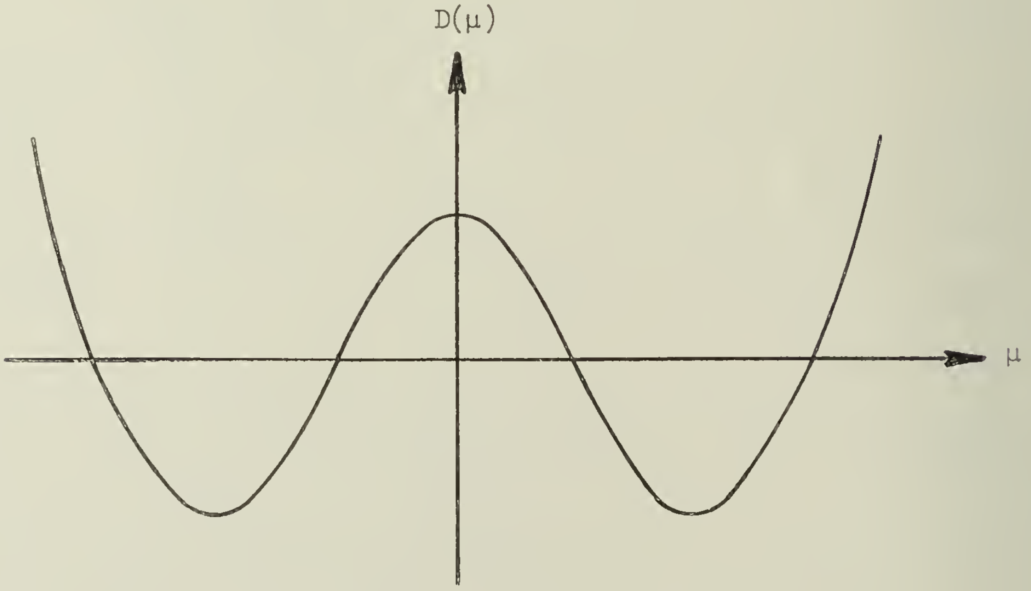

Figure 1

Theorem 2.5.2. Let ρ_μ denote the maximum of the moduli of the roots of $\varphi_\mu(\lambda)$. Then $\max_{\underline{\mu} \leq \mu \leq \bar{\mu}} \rho_\mu$ is minimal if and only if ω_1 and ω_2 are such that $D(\underline{\mu}) = D(\bar{\mu}) = 0$

Proof. The equation $\varphi_\mu(\lambda) = 0$ can be written

$$\frac{1}{\omega_1\omega_2}\,[\lambda - (1 - \omega_1)][\lambda - (1 - \omega_2)] = \mu^2\,\lambda\,.$$

Completing the square on the left hand side yields

$$\frac{1}{\omega_1\omega_2}\left[\lambda - \frac{(1 - \omega_1) + (1 - \omega_2)}{2}\right]^2 = \mu^2\,\lambda + \frac{(\omega_2 - \omega_1)^2}{4\omega_1\omega_2}\quad .$$

Some algebraic manipulation converts this to

$$\left\{\frac{1}{\sqrt{\omega_1\,\omega_2}}\,(\lambda - 1) + \frac{1}{2}\left[\sqrt{\frac{\omega_1}{\omega_2}} + \sqrt{\frac{\omega_2}{\omega_1}}\right]\right\}^2 = \mu^2\,\lambda + \frac{1}{4}\left[\sqrt{\frac{\omega_1}{\omega_2}} - \sqrt{\frac{\omega_2}{\omega_1}}\right]^2 .$$

It follows that the roots of $\varphi_\mu(\lambda) = 0$ are the abscissas of the points of intersection in the (λ,σ) plane of the parabola

$$\sigma_1^2 = \sigma_1^2(\mu) = \mu^2\,\lambda + \frac{1}{4}\left[\sqrt{\frac{\omega_1}{\omega_2}} - \sqrt{\frac{\omega_2}{\omega_1}}\right]^2$$

[illegible]
and the line

$$\sigma_2 = \frac{1}{\sqrt{\omega_1\omega_2}}\,(\lambda - 1) + \frac{1}{2}\left[\sqrt{\frac{\omega_1}{\omega_2}} + \sqrt{\frac{\omega_2}{\omega_1}}\right]$$

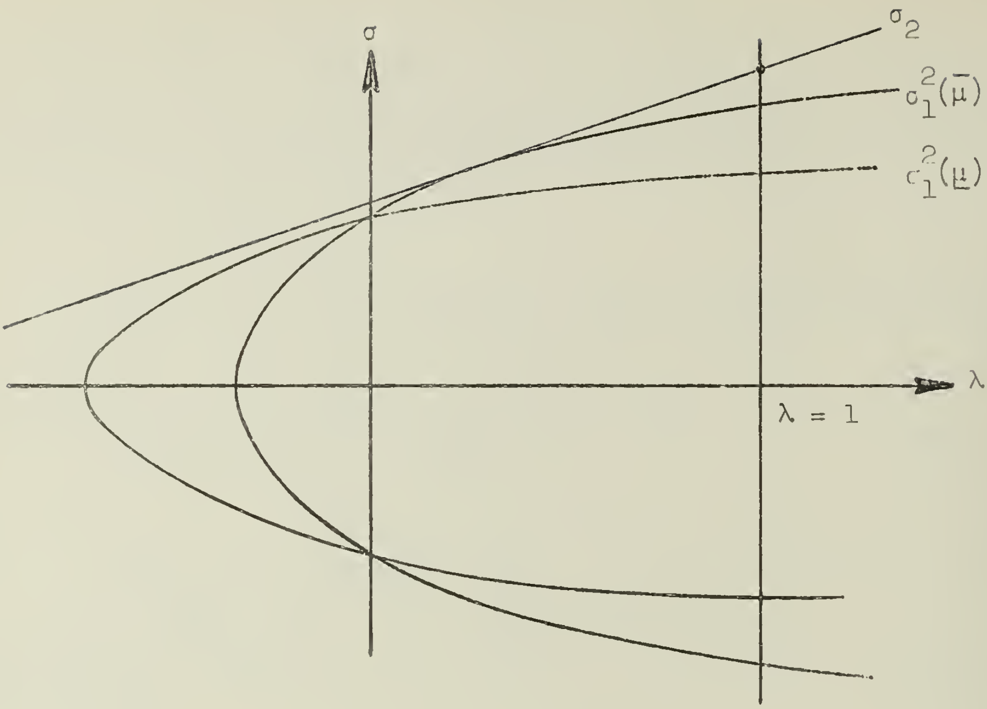

Figure 2

For $\frac{\omega_2}{\omega_1} = 1$, the vertices of the parabolas $\sigma_1^2(\mu)$ are at the origin. For $\frac{\omega_2}{\omega_1} \neq 1$, the parabolas intersect on the σ-axis, the vertices lie on the negative λ-axis, and the abscissas of the vertices increase toward zero as μ increases. For fixed $\frac{\omega_2}{\omega_1} \geq 1$ and $\omega_1\omega_2$ sufficiently small, the line σ_2 intersects the parabola $\sigma_1^2(\bar{\mu})$ in two distinct points. As $\omega_1\omega_2$ increases, the slope of the line decreases, and so $\rho_{\bar{\mu}}$ decreases until the line becomes tangent to the parabola. At this point the roots of $\varphi_{\bar{\mu}}(\lambda)$ are real and equal to $\sqrt{(1 - \omega_1)(1 - \omega_2)}$. If $\omega_1 > 1$ and $\omega_2 > 1$ (as will be shown to be the case of interest), then further increase of $\omega_1\omega_2$

causes the roots of $\varphi_{\bar\mu}(\lambda)$ to become complex and to increase in modulus. It follows then, that for $\frac{\omega_2}{\omega_1}$ fixed, the value of $\omega_1\omega_2$ which minimizes $\rho_{\bar\mu}$ is the value such that the line σ_2 is tangent to the parabola $\sigma_1^2(\bar\mu)$.

For $\frac{\omega_2}{\omega_1} > 1$ but sufficiently small, the line σ_2 does not intersect the parabola $\sigma_1^2(\underline\mu)$ when it is tangent to the parabola $\sigma_1^2(\bar\mu)$, and so the roots of $\varphi_{\underline\mu}(\lambda)$ are complex of modulus $\sqrt{(1 - \omega_1)(1 - \omega_2)}$. As $\frac{\omega_2}{\omega_1}$ increases the parabolas $\sigma_1^2(\bar\mu)$ and $\sigma_1^2(\underline\mu)$ move to the left and the distance between their vertices increases.

For $\frac{\omega_2}{\omega_1}$ sufficiently large, say $\frac{\omega_2}{\omega_1} = (\frac{\omega_2}{\omega_1})_b$, there exists a value of $\omega_1\omega_2$ which makes the line σ_2 simultaneously tangent to the parabolas $\sigma_1^2(\underline\mu)$ and $\sigma_1^2(\bar\mu)$. Hence, $D(\bar\mu) = D(\underline\mu) = 0$, and the roots of $\varphi_{\bar\mu}(\lambda)$ are both equal to $\sqrt{(1 - \omega_1)(1 - \omega_2)}$ while the roots of $\varphi_{\underline\mu}(\lambda)$ are both equal to $-\sqrt{(1 - \omega_1)(1 - \omega_2)}$. Moreover, it follows from lemma 2.3.5 that the roots of $\varphi_{\mu}(\lambda)$ for $\underline\mu < \mu < \bar\mu$ are complex of modulus $\sqrt{(1 - \omega_1)(1 - \omega_2)}$.

If $\frac{\omega_2}{\omega_1}$ increases further and $\omega_1\omega_2$ is adjusted so that the line remains tangent to the parabola $\sigma_1^2(\underline\mu)$, the point of tangency moves to the left because the parabola moves to the left. Hence min $\rho_{\underline\mu}$ regarded as a function of $\omega_1\omega_2$ increases, and the line ceases to intersect the parabola $\sigma_1^2(\bar\mu)$. It follows that $\max\limits_{\underline\mu \leq \mu \leq \bar\mu} \rho_{\mu}$ is minimal when ω_1 and ω_2 are such that the line is simultaneously tangent to both parabolas, which is the case if and only if $D(\bar\mu) = D(\underline\mu) = 0$.

This argument shows that if $\frac{\omega_2}{\omega_1} \geq 1$, then the pair (ω_1, ω_2) is optimal if and only if $D(\bar{\mu}) = D(\underline{\mu}) = 0$. This latter condition also characterizes an optimal pair such that $\frac{\omega_2}{\omega_1} \leq 1$, as can be shown by letting $\frac{\omega_2}{\omega_1}$ decrease from 1 in the above argument. Indeed, since $\varphi_\mu(\lambda)$ is a symmetric function of ω_1 and ω_2, it follows that (ω_1, ω_2) is an optimal pair if and only if (ω_2, ω_1) is also.

Theorem 2.3.3: Let $\underline{\omega}_b = \dfrac{2}{1 + \underline{\mu}\,\bar{\mu} + \sqrt{(1-\underline{\mu}^2)(1-\bar{\mu}^2)}}$,

$\bar{\omega}_b = \dfrac{2}{1 - \underline{\mu}\,\bar{\mu} + \sqrt{(1-\underline{\mu}^2)(1-\bar{\mu}^2)}}$, and $\rho_{2SOR} = \dfrac{\sqrt{1-\underline{\mu}^2} - \sqrt{1-\bar{\mu}^2}}{\sqrt{1-\underline{\mu}^2} + \sqrt{1-\bar{\mu}^2}}$. Then

(i) $\max\limits_{\underline{\mu} \leq \mu \leq \bar{\mu}} \rho_\mu = \rho_{2SOR} = \min\limits_{\omega_1, \omega_2} \left\{ \max\limits_{\underline{\mu} \leq \mu \leq \bar{\mu}} \rho_\mu \right\}$

if and only if either $\left\{ \begin{array}{l} \omega_1 = \underline{\omega}_b \\ \omega_2 = \bar{\omega}_b \end{array} \right\}$ or $\left\{ \begin{array}{l} \omega_1 = \bar{\omega}_b \\ \omega_2 = \underline{\omega}_b \end{array} \right\}$ holds.

(ii) If ω_b denotes the optimum extrapolation factor for SOR and $\rho_{SOR} = \rho(\mathcal{L}_{\omega_b})$, then $1 \leq \underline{\omega}_b \leq \omega_b \leq \bar{\omega}_b$ and $\rho_{2SOR} \leq \rho_{SOR}$.

(iii) $\underline{\omega}_b = \omega_b = \bar{\omega}_b$ and $\rho_{2SOR} = \rho_{SOR}$ if and only if $\underline{\mu} = 0$.

Proof:

(i) By Theorem 2.3.2, $\max\limits_{\underline{\mu} \leq \mu \leq \bar{\mu}} \rho_\mu$ is minimal if and only if $D(\underline{\mu}) = D(\bar{\mu}) = 0$. Since

$$D(\mu) = \omega_1\omega_2\mu^2\,[\omega_1\omega_2\mu^2 + 2(1-\omega_1) + 2(1-\omega_2)] + (\omega_2-\omega_1)^2$$

is quadratic in μ^2 with roots $\underline{\mu}^2$ and $\bar{\mu}^2$, we have

$$\frac{\underline{\mu} + \bar{\mu}}{2} = \frac{(\omega_1 + \omega_2 - 2)}{\omega_1 \omega_2} \tag{2.3.2}$$

and

$$\underline{\mu}^2 \bar{\mu}^2 = \frac{(\omega_2 - \omega_1)^2}{\omega_1^2 \omega_2^2} \quad . \tag{2.3.3}$$

Adding plus and minus the positive square root $\bar{\mu}\,\underline{\mu} = \dfrac{\omega_2 - \omega_1}{\omega_1 \omega_2}$ of (2.3.3) to (2.3.2) yields

$$(\bar{\mu} + \underline{\mu})^2 \omega_1\omega_2 - 4\omega_2 + 4 = 0$$

and

$$(\bar{\mu} - \underline{\mu})^2 \omega_1\omega_2 - 4\omega_1 + 4 = 0,$$

whose solutions are

$$\begin{Bmatrix} \omega_1 = \underline{\omega}_b \\ \omega_2 = \bar{\omega}_b \end{Bmatrix}$$

and

$$\begin{Bmatrix} \omega_1 = \dfrac{2}{1 + \underline{\mu}\,\bar{\mu} - \sqrt{(1 - \underline{\mu}^2)(1 - \bar{\mu}^2)}} \\ \\ \omega_2 = \dfrac{2}{1 - \underline{\mu}\,\bar{\mu} - \sqrt{(1 - \underline{\mu}^2)(1 - \bar{\mu}^2)}} \end{Bmatrix} \quad .$$

$|(1-\omega_1)(1-\omega_2)|$ is smaller for the solution

$$\left\{\begin{matrix} \omega_1 = \underline{\omega}_b \\ \omega_2 = \bar{\omega}_b \end{matrix}\right\} .$$

An analogous development using the negative square root $\bar{\mu}\,\underline{\mu} = \dfrac{\omega_1 - \omega_2}{\omega_1 \omega_2}$ of (2.3.3) yields the solution $\left\{\begin{matrix} \omega_1 = \bar{\omega}_b \\ \omega_2 = \underline{\omega}_b \end{matrix}\right\}$. Hence,

$$\min_{\omega_1, \omega_2} \left\{ \max_{\underline{\mu} \leq \mu \leq \bar{\mu}} \rho_\mu \right\} = \sqrt{(1 - \underline{\omega}_b)(1 - \bar{\omega}_b)} ,$$

and some algebra shows that $\sqrt{(1 - \underline{\omega}_b)(1 - \bar{\omega}_b)} = \rho_{2SOR}$

(ii) Since $0 \leq \underline{\mu} \leq \bar{\mu} < 1$, we can set $\underline{\mu} = \sin\alpha$ and $\bar{\mu} = \sin\beta$, where $0 \leq \alpha \leq \beta < \frac{\pi}{2}$. Then

$$\underline{\omega}_b = \frac{2}{1 + \sin\alpha \sin\beta + \cos\alpha \cos\beta} = \frac{2}{1 + \cos(\beta - \alpha)}$$

$$\bar{\omega}_b = \frac{2}{1 - \sin\alpha \sin\beta + \cos\alpha \cos\beta} = \frac{2}{1 + \cos(\beta + \alpha)}$$

and

$$\omega_b = \frac{2}{1 + \sqrt{1 - \bar{\mu}^2}} = \frac{2}{1 + \cos\beta} .$$

Since $0 \leq \alpha \leq \beta < \frac{\pi}{2}$, it follows that $\cos(\beta + \alpha) \leq \cos\beta \leq \cos(\beta - \alpha)$, and so $1 \leq \underline{\omega}_b \leq \omega_b \leq \bar{\omega}_b$.

If $\underline{\mu} = 0$, then

$$\rho_{2SOR} = \frac{1 - \sqrt{1 - \bar{\mu}^2}}{1 + \sqrt{1 - \bar{\mu}^2}} \equiv \rho_{SOR} .$$

But $\frac{d}{d\underline{\mu}} \rho_{2SOR} < 0$ for $\underline{\mu} > 0$, and therefore $\rho_{2SOR} \leq \rho_{SOR}$ for $\underline{\mu} \geq 0$.

(iii) This follows at once from the formulas which define the quantities involved.

It is of interest to trace the movement in the complex plane of the roots of $\varphi_\mu(\lambda)$ for fixed ω_1 and ω_2 satisfying $1 < \omega_1 \leq \omega_2$ as μ decreases from $+\infty$ to zero. It follows from (2.3.2) and (2.3.3) in the proof of the preceding theorem that each such pair (ω_1, ω_2) corresponds to some pair $(\underline{\mu}, \bar{\mu})$ satisfying $0 \leq \underline{\mu} \leq \bar{\mu}$.

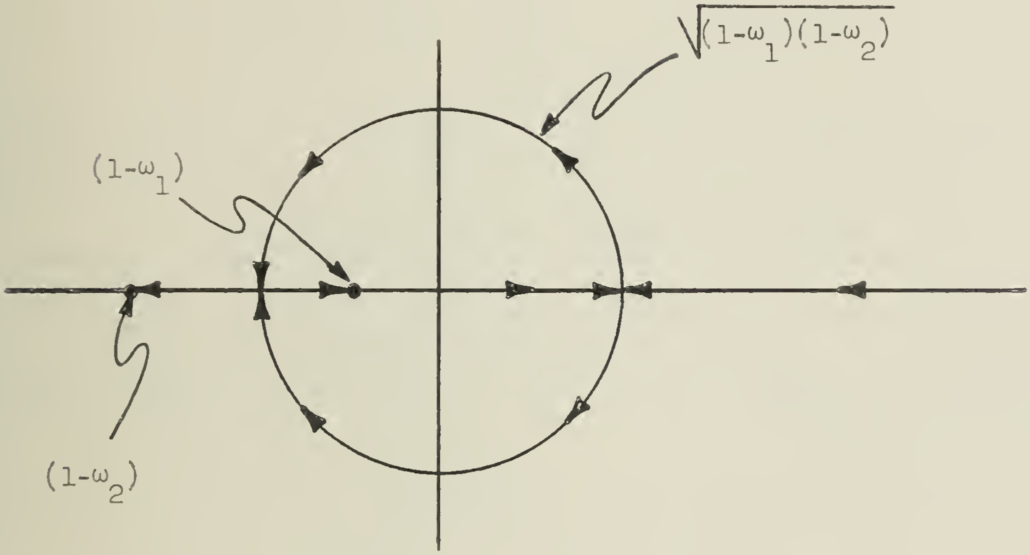

Figure 3

For $\mu = +\infty$, the roots of $\varphi_\mu(\lambda)$ are zero and $+\infty$. As μ decreases, the roots approach each other until for $\mu = \bar{\mu}$, they are equal to $\sqrt{(1 - \omega_1)(1 - \omega_2)}$. As μ decreases from $\bar{\mu}$ to $\underline{\mu}$, the roots move around the circle of radius $\sqrt{(1 - \omega_1)(1 - \omega_2)}$ in the complex plane, becoming equal to $-\sqrt{(1 - \omega_1)(1 - \omega_2)}$ when $\mu = \underline{\mu}$. As μ decreased from $\underline{\mu}$ to zero, the roots move apart on the real axis so that for $\mu = 0$, one root is $(1 - \omega_1)$ and the other is $(1 - \omega_2)$.

Theorem 2.3.4: Using the notation of lemmas 2.3.1 and 2.3.3, let

$$\underline{\mu} = \begin{Bmatrix} \mu_1 & \text{if } p = |r_o - r_e| \\ 0 & \text{if } p > |r_o - r_e| \end{Bmatrix},$$

$$\omega_1 = \begin{Bmatrix} \underline{\omega}_b & \text{if } r_o > r_e \\ \bar{\omega}_b & \text{if } r_o \leq r_e \end{Bmatrix}$$

$$\omega_2 = \begin{Bmatrix} \bar{\omega}_b & \text{if } r_o > r_e \\ \underline{\omega}_b & \text{if } r_o \leq r_e \end{Bmatrix} .$$

Then

$$\rho(\mathcal{L}_{\omega_1,\omega_2}) = \rho_{2SOR} = \min_{\omega_1,\ \omega_2} \rho(\mathcal{L}_{\omega_1,\omega_2}) .$$

Proof: If $p = |r_o - r_e|$, then by theorem 2.3.1 the characteristic polynomial of $\mathcal{L}_{\omega_1,\omega_2}$ is

$$\Phi(\lambda) = [(1 - \underline{\omega}_b) - \lambda]^p \prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda) .$$

By theorem 2.3.3, the maximum of the moduli of the roots of $\prod_{i=1}^{\ell} \varphi_{\mu_i}(\lambda)$ is minimal for either of the two choices for (ω_1,ω_2) specified in the theorem, and in fact the roots lie on a circle of radius ρ_{2SOR}. But the remaining roots of $\Phi(\lambda)$ lie inside this circle because $|1 - \underline{\omega}_b| < \sqrt{(1 - \underline{\omega}_b)(1 - \bar{\omega}_b)} = \rho_{2SOR}$, and the conclusion of the theorem follows.

If $p > |r_o - r_e|$, then both $(1 - \underline{\omega}_b)$ and $(1 - \bar{\omega}_b)$ are roots of $\Phi(\lambda)$. But these are the roots of $\varphi_\mu(\lambda)$ if $\mu = 0$, and so it follows that $\rho(\mathcal{L}_{\omega_1,\omega_2})$ is minimized by taking $\underline{\mu} = 0$.

2.4 General VSOR Extrapolation Matrices

Let A be partitioned so that each diagonal submatrix A_{ii} is square of order r_i. All of the iterations considered in this thesis for the solution of (2.1.1) are of the form

$$\vec{Z}_i^{\,m+1} = \Omega_i[B_{i,i-1}\vec{Z}_{i-1}^{\,m+1} + B_{i,i+1}\vec{Z}_{i+1}^{\,m}] + (I - \Omega_i)\vec{Z}_i^{\,m} + \Omega_i A_{ii}^{-1}\vec{K}_i, \; 1 \leq i \leq q \tag{2.4.1}$$

where each Ω_i is a non-singular square matrix of order r_i. For example, in the case of SOR or of the 2 factor VSOR of Section 2.3, each Ω_i is simply a scalar multiple of the identity matrix. The iterations to be investigated in Chapter 3 are also special cases of (2.4.1), although their implementation does not involve explicit multiplication by the Ω_i's.

If

$$\Omega = \begin{bmatrix} \Omega_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \Omega_q \end{bmatrix}$$

then the iteration matrix for (2.4.1) is

$$\mathcal{L}_\Omega = [I - \Omega L]^{-1}[\Omega U + (I - \Omega)]$$

where L + U = M, the block Jacobi matrix derived from A. If each Ω_i is non-singular and if $\vec{Z}^{\,m} = \vec{Z}$, the true solution of (2.1.1), then $\vec{Z}^{\,m+1} = \vec{Z}$ also.

In case each Ω_i is a full matrix, the implementation of (2.4.1) evidently entails considerably more work than extrapolation by a scalar. Suppose, for example, that $r_i = r$, $1 \leq i \leq q$. $\vec{Z}_i^{m+1}$ can be computed most economically (even if Ω_i is a scalar multiple of the identity) by first computing

$$\vec{Z}_i^{m+\frac{1}{2}} = B_{i,i-1}\vec{Z}_{i-1}^{m+1} + B_{i,i+1}\vec{Z}_{i+1}^{m} + A_{ii}^{-1}\vec{K}_i \tag{2.4.2}$$

and then extrapolating:

$$\vec{Z}_i^{m+1} = \Omega_i\,[\vec{Z}_i^{m+\frac{1}{2}} - \vec{Z}_i^{m}] + \vec{Z}_i^{m}\,. \tag{2.4.3}$$

(2.4.3) involves r^2 multiplications and $r(r+1)$ additions, and since $\vec{Z}_i$ contains r unknowns, it follows that the extrapolation procedure for (2.4.1) requires r multiplications and r+1 additions per unknown of the linear system (2.1.1). Extrapolation by a scalar, however, requires only one multiplication and two additions per unknown. The direct use, therefore, of full extrapolation matrices is computationally advantageous only if a large increase in the rate of convergence can be obtained over iterations using scalar extrapolation factors.

Such a large increase is indeed possible. In fact, for a properly chosen set $\{\Omega_i\}$ of extrapolation matrices, $\mathcal{L}_\Omega$ is nilpotent (i.e., $\rho(\mathcal{L}_\Omega) = 0$) so that the rate of convergence of (2.4.1) is infinite. The following theorem gives criteria for choosing such a set $\{\Omega_i\}$.

Theorem 2.4.1. $\mathfrak{L}_\Omega$ is nilpotent if $\Omega_1, \Omega_2, --- \Omega_q$ satisfy one of

(i) $\Omega_1 = I$ and $\Omega_i = [I - B_{i,i-1}\Omega_{i-1}B_{i-1,i}]^{-1}$, $2 \leq i \leq q$

(ii) $\Omega_q = I$ and $\Omega_i = [I - B_{i,i+1}\Omega_{i+1}B_{i+1,i}]^{-1}$, $1 \leq i \leq q-1$

(iii) $\Omega_1 = I$, $\Omega_q = I$, and for some m such that $1 < m < q$;

$\Omega_i = [I - B_{i,i-1}\Omega_{i-1}B_{i-1,i}]^{-1}$, $2 \leq i < m$;

$\Omega_i = [I - B_{i,i+1}\Omega_{i+1}B_{i+1,i}]^{-1}$, $m+1 \leq i \leq q$;

$\Omega_m = [I - B_{m,m-1}\Omega_{m-1}B_{m-1,m} - B_{m,m+1}\Omega_{m+1}B_{m+1,m}]^{-1}$

Proof: Suppose that

$$\vec{X} = [\vec{X}_1^T \mid \vec{X}_2^T \mid --- \mid \vec{X}_q^T]^T$$

is an eigenvector of $\mathfrak{L}_\Omega$ belonging to the eigenvalue λ. It follows from (2.4.1) that for $1 \leq i \leq q$, $\lambda\vec{X}_i = \Omega_i[\lambda B_{i,i-1}\vec{X}_{i-1} + B_{i,i+1}\vec{X}_{i+1}] + (I-\Omega_i)\vec{X}_i$, and hence $G\vec{X} = \vec{0}$, where G is

$$\begin{bmatrix} [(I_1-\Omega_1)-\lambda I_1] & \Omega_1 B_{12} & & \bigcirc \\ \lambda\Omega_2 B_{21} & \ddots & \ddots & \\ & \ddots & \ddots & \Omega_{q-1}B_{q-1,q} \\ \bigcirc & & \lambda\Omega_q \quad B_{q,q-1} & [(I_q-\Omega_q)-\lambda I_q] \end{bmatrix}$$

Here I_i is the identity matrix of order r_i.

There exists a non-trivial solution of $G\vec{X} = \vec{0}$ if and only if $\det(G) = 0$, and since $\det(G)$ is a polynomial of degree n in λ, it follows that its roots are precisely the eigenvalues of $\mathcal{L}_{\Omega}$. We now show that if the Ω_i's satisfy (i), (ii), or (iii), then $\det(G) = \pm \lambda^n$.

If (i) is satisfied, then the block in the (1,1) position of G is $-\lambda I_1$. We now perform the block elementary row operation of adding $\Omega_2 B_{21}$ times the first r_1 rows of G to the next r_2 rows. That is, we add $\Omega_2 B_{21}(-\lambda I_1)$ to $\lambda\Omega_2 B_{21}$ and $\Omega_2 B_{21}\Omega_1 B_{12}$ to $[(I - \Omega_2) - \lambda I_2]$. This operation leaves det(G) unaltered since it corresponds to premultiplication of G by

$$\begin{bmatrix} I & & & \\ B_{21} & I & & \bigcirc \\ & & \ddots & \\ \bigcirc & & & I \end{bmatrix}$$

whose determinant is obviously 1.

The block in the (2,1) position is now null, while the block in the (2,2) position is $[\Omega_2 B_{21}\Omega_1 B_{12} + (I - \Omega_2)] - \lambda I_2$. The quantity in square brackets vanishes because Ω_2 satisfies the recursion relation (i), and so the diagonal block in the (2,2) position is $-\lambda I_2$. Hence, we repeat the procedure, adding $\Omega_3 B_{32}$ times the second r_2 rows to the next r_3 rows. The sub-diagonal block in the (3,2) position vanishes, while the diagonal block in the (3,3) position becomes $-\lambda I_3$ since Ω_3 satisfies (i). By continuing in this way, all sub-diagonal blocks of G are made to vanish (i.e., G is made block upper triangular), and the i-th diagonal block becomes $-\lambda I_i$. This procedure leaves det(G) unaltered, and $\det(G) = \pm \lambda^n$. Hence $\mathcal{L}_{\Omega}$ is nilpotent.

In case the Ω_i's satisfy (ii), we proceed by block column operations starting from the right to reduce G to the same upper triangular form.

If (iii) is satisfied, we proceed by block row operations starting from above to eliminate the sub-diagonal blocks in positions (2,1), (3,2), --- (m,m-1) and by block column operations starting from the right to eliminate the remaining sub-diagonal blocks. Again the i-th diagonal block becomes $-\lambda I_i$. Hence, if one of (i), (ii), or (iii) is satisfied, $\det(G) = \pm \lambda^n$, and so $\mathcal{L}_\Omega$ is nilpotent in all cases.

Q.E.D.

If $\mathcal{L}_\Omega$ is nilpotent, then the error in the successive iterates eventually becomes zero in the absence of rounding error, and so the exact solution to (2.1.1) is obtained after a finite number of iterations. By the Cayley-Hamilton theorem at most n iterations are required, and in fact it can be shown that q iterations suffice. Hence, the iteration (2.4.1) with the Ω_i's chosen according to the criteria of of theorem 2.4.1 is more properly classed as a direct method than as an iterative one, and direct methods are not the principal concern of this thesis. In Chapter 3, however, we will use the criteria of theorem 2.4.1 not to obtain the Ω_i's themselves, but to obtain certain <u>scalar</u> acceleration parameters used in the actual iteration.

2.5 Extrapolation Matrices When the Block Matrices Have a Common Basis of Eigenvectors

Consider now the linear system (2.1.1) arising from the discretization of the Dirichlet problem on a rectangular domain R for a self-adjoint, elliptic partial differential equation of the special form

$$- \frac{\partial}{\partial x} (f_1(x) \frac{\partial u(x,y)}{\partial x}) - \frac{\partial}{\partial y} (f_2(y) \frac{\partial u(x,y)}{\partial y})$$
$$+ (\sigma_1(x) + \sigma_2(y))u(x,y) = g(x,y), \tag{2.5.1}$$

where $f_1(x)$, $f_2(y)$, $\sigma_1(x)$, $\sigma_2(y)$ are continuous in the closure of R and satisfy $f_1(x) > 0$, $f_2(y) > 0$, $\sigma_1(x) > 0$, $\sigma_2(y) > 0$. Let a uniform mesh be imposed on R, and suppose that the mesh points are numbered either by rows or by columns starting from the upper left. Thus, the mesh points (and hence the unknowns in (2.1.1)) are grouped into q blocks of r elements each. It follows that for $1 \leq i \leq q$, the blocks $B_{i,i-1}$ and $B_{i,i+1}$ of the block Jacobi matrix derived from A are square of order r and all have a common basis of eigenvectors (i.e., a set of r linearly independent eigenvectors). In this case the matrices $\{\Omega_i\}$ can be chosen in various ways so that they share this common basis of eigenvectors, and consequently a useful factorization of the characteristic polynomial of $\mathfrak{L}_\Omega$ is possible. In Chapter 3 we will investigate some iterations which are equivalent to the use of extrapolation matrices having the same eigenvectors as the matrices $B_{i,i-1}$ and $B_{i,i+1}$. The matrices $\{\Omega_i\}$ will then be determined by specifying certain of their eigenvalues.

Let $\beta^j_{i,i-1}$, $\beta^j_{i,i+1}$, and ω^j_i denote the eigenvalues of $B_{i,i-1}$, $B_{i,i+1}$ and Ω_i respectively to which the (r-dimensional) eigenvector $\vec{\xi}^j$ belongs, and consider the scalar iteration

$$x^{m+1}_i = \omega^j_i (\beta^j_{i,i-1} x^{m+1}_{i-1} + \beta^j_{i,i+1} x^m) + (1 - \omega^j_i) x^m_i, \tag{2.5.2}$$

$$1 \leq i \leq q.$$

Let $\mathcal{L}_\omega^j$ denote the iteration matrix for (2.5.2), and let $\Phi^j(\lambda)$ denote the characteristic polynomial of $\mathcal{L}_\omega^j$. We then have the following decomposition theorem, to be used in Chapter 3.

<u>Theorem 2.5.1</u>: If the matrices $B_{i,i-1}$, $B_{i,i+1}$, and Ω_i have a common basis of eigenvectors for $1 \leq i \leq q$, then the characteristic polynomial of $\mathcal{L}_\Omega$ is $\Phi(\lambda) = \prod_{j=1}^{r} \Phi^j(\lambda)$.

<u>Proof</u>: If $\vec{x} = [x_1, x_2, ---, x_q]^T$ is an eigenvector of $\mathcal{L}_\omega^j$ belonging to the eigenvalue λ, then it follows from (2.5.2) that

$$\lambda x_i = \omega_i^j(\lambda\beta_{i,i-1}x_{i-1} + \beta_{i,i+1}x_{i+1}) + (1-\omega_i^j)x_i,\ 1 \leq i \leq q.$$

Hence $G^j\,\vec{X} = \vec{0}$, where

$$G^j = \begin{bmatrix} [(1-\omega_1^j)-\lambda] & \omega_1^j\beta_{12}^j & & \bigcirc \\ \lambda\omega_2^j\beta_{21}^j & \ddots & \ddots & \\ & \ddots & \ddots & \omega_{q-1}^j\beta_{q-1,q}^j \\ \bigcirc & & \lambda\omega_q^j\ \beta_{q,q-1}^j & [(1-\omega_q^j)-\lambda] \end{bmatrix}$$

It follows that λ is an eigenvalue of $\mathcal{L}_\omega^j$ if and only if λ is a root of $\det(G^j) = 0$, and therefore $\Phi^j(\lambda) = \det(G^j)$.

Let H be the square matrix of order r whose j-th column is $\vec{\xi}^j$, and let $\bar{H} = \sum^q \oplus H$, where $\oplus$ denotes the direct sum. The linear independence of $\vec{\xi}^1, \vec{\xi}^2, ---, \vec{\xi}^r$ implies that H is non-singular and so is $\bar{H}$. $\Phi(\lambda) = \det(G)$, where G was defined in the proof of theorem 2.4.1, and it follows that $\Phi(\lambda) = \det(\bar{H}^{-1}G\bar{H})$. G is block tridiagonal, and each non-null block of G is square of order r. Hence, the effect of transformation of G by $\bar{H}$ is to transform each non-null block of G by H, and diagonalize it. Thus,

$$H^{-1}\Omega_i B_{i,i-1} H = \begin{bmatrix} \omega_i^1\beta_{i,i-1}^1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \omega_i^r\beta_{i,i-1}^r \end{bmatrix}$$

$$H^{-1}\Omega_i B_{i,i+1} H = \begin{bmatrix} \omega_i^1\beta_{i,i+1}^1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \omega_i^r\beta_{i,i+1}^r \end{bmatrix}$$

and

$$H^{-1}[(I-\Omega_i)-\lambda I_i]H = \begin{bmatrix} [(1-\omega_i^1)-\lambda] & & \bigcirc \\ & \ddots & \\ \bigcirc & & [(1-\omega_i^r)-\lambda] \end{bmatrix}$$

Hence, $\bar{H}^{-1}\bar{G}\bar{H}$ is an shown in Figure 4. .

Let σ be the permutation which selects the first element of the first block first, the first element of the second block second, and so on through the first elements of the various blocks, then likewise through the second elements of the various blocks, etc. If P_σ is the permutation matrix corresponding to σ, then $\Phi(\lambda) = \det(P_\sigma\bar{H}^{-1}\bar{G}\bar{H}P_\sigma^T)$. But $P_\sigma\bar{H}^{-1}\bar{G}\bar{H}P_\sigma^T = \sum_{j=1}^{q} \oplus G^j$, and hence $\Phi(\lambda) = \prod_{j=1}^{q} \det(G^j) = \prod_{j=1}^{q} \Phi^j(\lambda)$. Q.E.D.

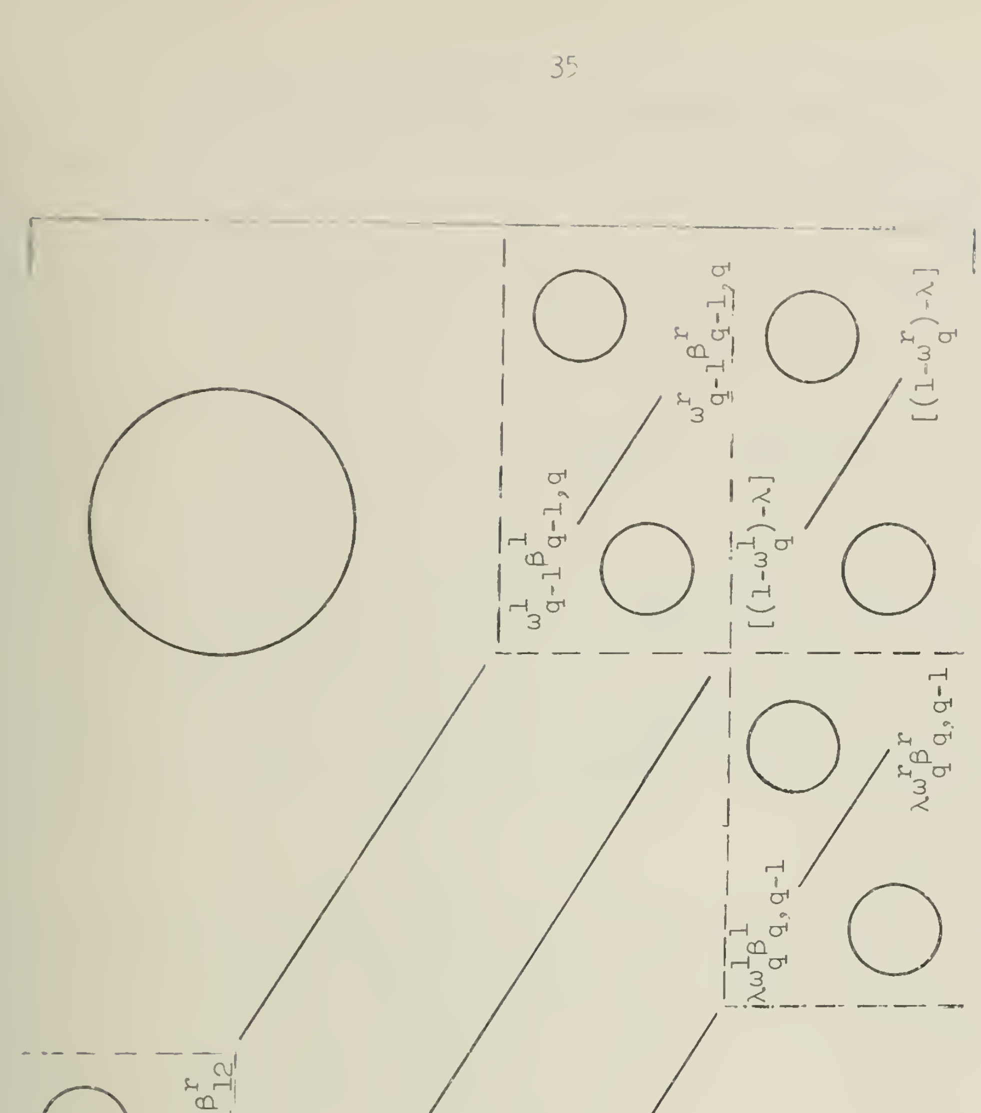
$[(1-\omega_1^l)-\lambda]$
$\omega_1^l\beta_{12}^l$
$[(1-\omega_1^r)-\lambda]$
$\omega_1^r\beta_{12}^r$
$\lambda\omega_2^l\beta_{21}^l$
$\lambda\omega_2^r\beta_{21}^r$
$\omega_{q-1}^l\beta_{q-1,q}^l$
$\omega_{q-1}^r\beta_{q-1,q}^r$
$[(1-\omega_q^l)-\lambda]$
$[(1-\omega_q^r)-\lambda]$
$\lambda\omega_q^l\beta_{q,q-1}^l$
$\lambda\omega_q^r\beta_{q,q-1}^r$

Figure 4

3. SEQUENTIAL EXTRAPOLATED IMPLICIT METHODS

3.1 A Model Problem

The usual discretization of the Dirichlet problem for Poisson's equation

$$\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2} = f(x,y) \qquad (3.1.1)$$

on a rectangle R using a five point star leads to a linear system whose coefficient matrix A is block tri-diagonal with tri-diagonal diagonal blocks. Let a uniform r x q mesh be imposed on the interior of R, and suppose that the mesh points are numbered by columns starting from the upper left. Then

$$A = \begin{bmatrix} D & -I & & \\ -I & \ddots & \ddots & \\ & \ddots & \ddots & -I \\ & & -I & D \end{bmatrix},$$

where I is the identity matrix of order r, D is of order r, and

$$D = \begin{bmatrix} 4 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 4 \end{bmatrix}.$$

The resulting linear system (2.1.1) will be referred to in what follows as the *model problem*.

For the model problem, each non-null block of the block Jacobi matrix derived from A is simply D^{-1}, and the block Gauss-Seidel iteration for the solution of (2.1.1) is

$$\vec{Z}_i^{\,m+1} = D^{-1}[\vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m} + \vec{K}_i], \quad 1 \le i \le q, \tag{3.1.2}$$

or

$$D\vec{Z}_i^{\,m+1} = \vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m} + \vec{K}_i, \quad 1 \le i \le q. \tag{3.1.3}$$

The block SOR iteration obtained by extrapolating (3.1.2) is frequently called SLOR (successive line over-relaxation) since it consists of updating simultaneously the unknowns corresponding to an entire vertical line of mesh points

3.2 The Sequential Extrapolated Implicit Method (SEI)

If $\vec{Z}^{\,m+1} = \vec{Z}^{\,m} = \vec{Z}$, the exact solution of the model problem, then for any scalar s,

$$(D + sI)\vec{Z}_i^{\,m+1} = \vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m} + sI\vec{Z}_i^{\,m} + \vec{K}_i\ , \quad 1 \le i \le q, \tag{3.2.1}$$

or

$$\vec{Z}_i^{\,m+1} = (D + sI)^{-1}\ [\vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m} + sI\vec{Z}_i^{\,m} + \vec{K}_i], \tag{3.2.2}$$

$$1 \le i \le q.$$

The iteration (3.2.2) can be written as

$$\vec{Z}_i^{\,m+1} = \Omega D^{-1}\ [\vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m} + \vec{K}_i] + (I - \Omega)\vec{Z}_i^{\,m} \tag{3.2.3}$$

$$1 \le i \le q\ ,$$

where $\Omega = (D + sI)^{-1} D$. That is, (3.2.2) is the VSOR iteration (2.4.1) with $\Omega_i = (D + sI)^{-1} D$, $1 \leq i \leq q$. We denote the iteration matrix for (3.2.2) by S_ω.

The most efficient way to perform SLOR is to use (3.1.2) to compute $\vec{Z}_i^{m+\frac{1}{2}}$ and then to compute

$$\vec{Z}_i^{m+1} = \omega[\vec{Z}_i^{m+\frac{1}{2}} - \vec{Z}_i^m] + \vec{Z}_i^m, \; 1 \leq i \leq q . \qquad (3.2.4)$$

Hence, the additional work required to perform SLOR over that required for Gauss-Seidel iteration is one multiplication and two additions per unknown.

Iteration (3.2.2) requires that $(D + sI)^{-1}$ be computed instead of D^{-1}, but both of these matrix inversions require the same amount of work since the diagonal of D is non-zero. Therefore the only additional work required to perform SEI over that for Gauss-Seidel iteration is one multiplication and one addition per unknown in order to compute the vector on the right hand side of (3.2.1). Hence SEI requires one less addition per unknown per iteration than SLOR.

Any eigenvector of D is also an eigenvector of Ω, and since the eigenvectors of D form a basis, the converse also holds. Hence, the decomposition theorem 2.5.1 applies.

Let d be an eigenvalue of D and $\vec{\xi}$ an eigenvector belonging to d. If β denotes the corresponding eigenvalue of D^{-1}, then $\beta = \frac{1}{d}$, and the eigenvalue of $\Omega = (D + sI)^{-1}D$ to which $\vec{\xi}$ belongs is $\frac{1}{1+s\beta}$, which we denote by $\omega(\beta, s)$.

Let $\mathcal{L}_\omega^\beta$ denote the matrix for the (scalar) iteration

$$x_i^{m+1} = \omega(\beta)\beta(x_{i-1}^{m+1} + x_{i+1}^{m}) + (1-\omega(\beta))x_i^m, \quad 1 \leq i \leq q, \qquad (3.2.5)$$

and let $\psi^\beta(\lambda)$ be the characteristic polynomial of $\mathcal{L}_\omega^\beta$. It follows from theorem 2.5.1 that if $\omega(\beta) = \omega(\beta,s)$ for all β, then the characteristic polynomial of S_ω is $\psi(\lambda) = \prod_\beta \psi^\beta(\lambda)$, where the product is taken over all eigenvalues of D^{-1}.

Let $\rho_{\omega(\beta)}^\beta$ denote the maximum of the moduli of the zeroes of $\psi^\beta(\lambda)$ and let $\omega_b(\beta)$ denote the value of $\omega(\beta)$ which minimizes $\rho_{\omega(\beta)}^\beta$. It follows from the theory of SOR that for $\omega(\beta) \geq \omega_b(\beta)$, the zeroes of $\psi^\beta(\lambda)$ are complex of modulus $(\omega(\beta)-1)$.

For fixed s, let $\rho(S_\omega)$ denote the spectral radius of S_ω. If $\underline{\beta}$ and $\bar{\beta}$ are the smallest and largest eigenvalues respectively of D^{-1}, then $\rho(S_\omega) = \max\limits_{\underline{\beta} \leq \beta \leq \bar{\beta}} \rho_{\omega(\beta,s)}^\beta$, and we let $\rho_{SEI} = \min\limits_s \rho(S_\omega)$.

The eigenvalues of D^{-1} are

$$\beta_j = \frac{1}{4 - 2\cos\left(\frac{j\pi}{r+1}\right)}, \quad 1 \leq j \leq r,$$

and so

$$\frac{1}{6} < \underline{\beta} \leq \beta_j \leq \bar{\beta} < \frac{1}{2}.$$

Let

$$\bar{A} = \begin{bmatrix} 0 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{bmatrix}$$

be square of order q. The eigenvalues of $\bar{A}$ are $\alpha_i = 2\cos\frac{i\pi}{q+1}$, $1 \leq i \leq q$, and since the block Jacobi matrix derived from A is $M = \bar{A} \otimes D^{-1}$ where $\otimes$ denotes the tensor product, it follows that the eigenvalues of M are $\mu_{ij} = \alpha_i\beta_j$, $1 \leq i \leq q$, $1 \leq j \leq r$. Moreover, the theory of SOR applied to the iteration (3.2.5) shows that

$$\omega_b(\beta) = \frac{2}{1 + \sqrt{1-\bar{\alpha}^2\beta^2}}, \tag{3.2.6}$$

where $\bar{\alpha} = \max_i \{\alpha_i\}$. Since the largest positive eigenvalue of M is $\bar{\mu} = \bar{\alpha}\,\bar{\beta}$, we have

$$\omega_b = \omega_b(\bar{\beta}) = \frac{2}{1 + \sqrt{1-\bar{\alpha}^2\bar{\beta}^2}}.$$

The VSOR iteration (2.4.1) becomes SOR if $\Omega_i = \omega_b I$, $1 \leq i \leq q$. In this case, theorem 2.5.1 shows that the characteristic polynomial of $\mathfrak{L}_\Omega = \mathfrak{L}_{\omega_b}$ is $\Phi(\lambda) = \prod_\beta \Phi^\beta(\lambda)$, where $\Phi^\beta(\lambda)$ is the characteristic polynomial of $\mathfrak{L}_\omega^\beta$ with $\omega(\beta) = \omega_b$. The maximum of the moduli of the zeroes of $\Phi^\beta(\lambda)$ is $\rho_{\omega_b}^\beta$.

Lemma 3.2.1: Let $s_b(\beta) = \frac{1}{\beta}\left[\frac{1}{\omega_b(\beta)} - 1\right]$, and denote $s_b(\bar{\beta})$ by s_b. Then for $\beta < \bar{\beta}$, $\omega_b(\beta) < \omega(\beta, s_b) < \omega_b(\bar{\beta}) = \omega_b$.

Proof: $s_b(\beta) < 0$ since $\omega_b(\beta) > 1$, and so $1 + s_b\bar{\beta} < 1 + s_b\beta$ if $0 < \beta < \bar{\beta}$. Hence, $\omega(\beta, s_b) < \omega_b(\bar{\beta})$. Substituting (3.2.6) into the formula for $s_b(\beta)$ and differentiating shows that $\frac{d}{d\beta} s_b(\beta) < 0$, and hence $s_b < s_b(\beta)$ for $\beta < \bar{\beta}$. Then

$$\frac{1}{\omega(\beta, s_b)} = 1 + s_b\beta < 1 + s_b(\beta)\beta = \frac{1}{\omega_b(\beta)},$$

and so $\omega_b(\beta) < \omega(\beta, s_b)$.

Q.E.D.

Theorem 3.2.1: If $s = s_b$, then:

(i) $\psi^{\bar\beta}(\lambda) = \Phi^{\bar\beta}(\lambda)$, and hence $\rho^{\bar\beta}_{\omega_b}(\bar\beta) = \rho_{SOR}$

(ii) For $\beta < \bar\beta$, $\rho^{\beta}_{\omega_b}(\beta) < \rho^{\beta}_{\omega(\beta,s_b)} < \rho_{SOR}$

(iii) $\lim\limits_{\beta\to 0} \rho^{\beta}_{\omega(\beta,s)} = 0$ for any s.

Proof:

(i) By definition, s_b is that value of s such that $\omega(\bar\beta,s) = \omega_b(\bar\beta)$. But $\omega_b(\bar\beta) = \omega_b$, and so $\psi^{\bar\beta}(\lambda) = \Phi^{\bar\beta}(\lambda)$. Since the moduli of the roots of $\Phi^{\beta}(\lambda)$ are equal to ρ_{SOR} for all $\beta \le \bar\beta$, it follows that $\rho^{\bar\beta}_{\omega_b}(\bar\beta) = \rho_{SOR}$.

(ii) It follows from lemma 3.2.1 and the theory of SOR that the roots of $\psi^{\beta}(\lambda)$ are complex of modulus $(\omega(\beta,s_b)-1)$ for $\beta \le \bar\beta$, and $(\omega_b(\beta)-1)) < (\omega(\beta,s_b)-1) < (\omega_b(\bar\beta)-1)$. The assertion (ii) then follows from the definitions of the quantities involved.

(iii) $\lim\limits_{\beta\to 0} \omega(\beta,s) = 1$, and since $\rho^{\beta}_{\omega(\beta,s)} = (\omega(\beta,s)-1)$, we have $\lim\limits_{\beta\to 0} \rho^{\beta}_{\omega(\beta,s)} = 0$.

Q.E.D.

Corrollary 3.2.1: $\rho_{SEI} = \rho_{SOR}$.

Proof: The theory of SOR applied to $\mathcal{L}^{\bar\beta}_{\omega}$ shows that $\rho^{\bar\beta}_{\omega(\bar\beta,s)}$ is minimal if and only if $\omega(\bar\beta,s) = \omega_b(\bar\beta)$, which is the case if and only if $s = s_b$. It then follows from part (ii) of theorem 3.2.1 that $\max\limits_{\underline{\beta} \le \beta \le \bar\beta} \rho^{\beta}_{\omega(\beta,s)}$ is minimal if and only if $s = s_b$, and part (i) of the theorem shows that $\rho_{SEI} = \rho_{SOR}$.

Q.E.D.

Let $\vec{\xi}^{\,j}$ normalized to have (Euclidean) length 1 be the eigenvector of D^{-1} belonging to β_j, $1 \leq j \leq r$. The vectors $\{\vec{\xi}^{\,j}\}_1^r$ form an ortho-normal basis for r-space since D^{-1} is symmetric, and indeed

$$\vec{\xi}^{\,j} = [\sin(\frac{j\Pi}{r+1}), \sin(\frac{2j\Pi}{r+1}), ---, \sin(\frac{rj\Pi}{r+1})]^T .$$

Let $\vec{e}^{\,i}$ denote the q-dimensional column vector having 1 in the i-th position and 0's elsewhere. The vectors $\{\vec{e}^{\,i}\}_1^q$ form an ortho-normal basis for q-space, and therefore the vectors $\{\vec{e}^{\,i} \otimes \vec{\xi}^{\,j}\}$, $1 \leq i \leq q$, $1 \leq j \leq r$, form an ortho-normal basis for qr-space, where $\otimes$ denotes the tensor product.

The eigenvector of D^{-1} corresponding to $\bar{\beta}$ is $\vec{\xi}^{\,1}$, and it follows from part (i) of theorem 3.2.1 that if the initial error $\vec{E}^{\,0}$ lies in the subspace spanned by $\{\vec{e}^{\,i} \otimes \vec{\xi}^{\,1}\}_1^q$ then the error $\vec{E}^{\,m}$ in the m-th iterate for SEI is identical to the error in the m-th iterate for SLOR, $\forall m$. Because of part (ii) of the same theorem, however, it is to be expected that if $\vec{E}^{\,0}$ does not lie entirely in this subspace, then a fixed number of iterations with SEI will produce a greater reduction in the ℓ_2 norm of the error than the same number of iterations with SLOR. Numerical results which support this expectation are presented in Section 3.4.

3.3 Cyclic Chebyshev SEI

The cyclic Chebyshev semi-iterative method [6; 11, Chap. 5] for the solution of the model problem is

$$\vec{Z}_i^{\,m+1} = \omega^{2m+1} D^{-1}[\vec{Z}_{i-1}^{\,m} + \vec{Z}_{i+1}^{\,m} + \vec{K}_i] + (1-\omega^{2m+1})\vec{Z}_i^{\,m}, \quad (3.3.1a)$$

$$i = 1, 3, 5, --- ,$$

$$\vec{Z}_{i}^{\,m+1} = \omega^{2m+2}\, D^{-1}[\vec{Z}_{i-1}^{\,m+1} + \vec{Z}_{i+1}^{\,m+1} + \vec{K}_{i}] + (1-\omega^{2m+2})\vec{Z}_{i}^{\,m}, \qquad (3.3.1b)$$

$$i = 2, 4, 6, \text{---}\ ,$$

where the sequence $\{\omega^m\}$ satisfies

$$\omega^1 = 1, \quad \omega^2 = \frac{1}{1-\frac{\bar{\mu}^2}{2}}\ ,$$

$$\omega^{m+1} = \frac{1}{1-\frac{\bar{\mu}^2}{4}\,\omega^m}\ , \quad m \geq 2, \qquad (3.3.2)$$

and $\qquad \bar{\mu} = \rho(M)$.

In what follows (3.3.1) with the sequence $\{\omega^m\}$ satisfying (3.3.2) will be called *cyclic Chebyshev SLOR*, while (3.3.1) with $\omega^m = \omega_b$, $\forall m$, will be called *SLOR with the odd-even ordering*. Iteration (2.2.3) with $\omega = \omega_b$ will be called *SLOR with the natural ordering*.

If the sequence $\{\omega^m\}$ satisfies (3.3.2), then $\lim_{m\to\infty} \omega^m = \omega_b$, and it follows that the *asymptotic* rate of convergence of cyclic Chebyshev SLOR is equal to that of SLOR. However, if $\vec{E}_c^{\,m}$ denotes the error in the m-th iterate for cyclic Chebyshev SLOR and if $\vec{E}^{\,m}$ denotes the error in the m-th iterate for SLOR (with any consistent ordering), then Golub and Varga [6] have shown that

$$\max_{\vec{Z}^0} ||\vec{E}_c^{\,m}|| < \max_{\vec{Z}^0} ||\vec{E}^{\,m}||, \quad m \leq 2,$$

where $||\quad||$ denotes the ℓ_2 norm. Moreover, the sequence $\{||\vec{E}_c^{\,m}||\}$ is strictly decreasing, whereas the sequence $\{||\vec{E}^{\,m}||\}$ may increase initially.

Let the sequence $\{\omega^m\}$ satisfy (3.3.2), and define a sequence $\{s^m\}$ by

$$s^m = \frac{1}{\bar{\beta}}\left[\frac{1}{\omega^m} - 1\right], \quad m \geq 1 . \qquad (3.3.3)$$

For $0 \leq \beta \leq \bar{\beta}$, define the sequence $\{\omega^m(\beta, s^m)\}$ by

$$\omega^m(\beta, s^m) = \frac{1}{1 + s^m\beta}, \quad m \geq 1. \qquad (3.3.4)$$

Then

$$\omega^m(\bar{\beta}, s^m) = \omega^m, \ \forall m.$$

The iteration

$$\vec{Z}_i^{m+1} = (D + s^{2m+1}I)^{-1}\,[\vec{Z}_{i-1}^{m} + \vec{Z}_{i+1}^{m} + s^{2m+1}\,\vec{Z}_i^{m} + \vec{K}_i],$$

$$i = 1, 3, 5, --- , \qquad (3.3.5a)$$

$$\vec{Z}_i^{m+1} = (D + s^{2m+2}I)^{-1}\,[\vec{Z}_{i-1}^{m+1} + \vec{Z}_{i+1}^{m+1} + s^{2m+2}\,\vec{Z}_i^{m} + \vec{K}_i]$$

$$i = 2, 4, 6, --- \qquad (3.3.5b)$$

for the solution of the model problem will be called cyclic Chebyshev SEI. Iteration (3.3.5) with s^m replaced by s_b, $\forall m$, will be called SEI with the odd-even ordering, while (3.2.2) with $s = s_b$ will be called SEI with the natural ordering.

The remarks in Section 3.2 concerning the convergence properties of SEI relative to those of SLOR apply also to the convergence properties of cyclic Chebyshev SEI relative to those of cyclic Chebyshev SLOR.

3.4 Numerical Results

Numerical experiments were performed to compare the actual rates of convergence of SLOR and SEI with both the natural and the odd-even ordering, cyclic Chebyshev SLOR, and cyclic Chebyshev SEI. The test problem selected was the model problem with $f(x,y) \equiv 0$ (i.e., Laplace's equation) and homogeneous boundary conditions. The rectangle R was taken to be a square containing N mesh points on a side. For N = 10, 20, 30, and 40 the number of iterations with each method needed to reduce the ℓ_2 norm of the error below 1 was determined.

Starting vectors were constructed as follows. Let $\vec{Q}$ be the N-dimensional vector whose j-th component is $(-1)^{(j+1)}$, and let $\vec{J}$ be the N-dimensional vector whose j-th component is j. Let $\vec{\eta} = \sum_{j=1}^{N} \vec{\xi}^{\,j}$, where $\vec{\xi}^{\,1}, \vec{\xi}^{\,2}, ---, \vec{\xi}^{\,N}$ are the eigenvectors of D^{-1}. Experiments were performed using each of the three starting vectors $\vec{Z}^{\,0} = \vec{J} \otimes \vec{Q}$, $\vec{Z}^{\,0} = 10 \vec{Q} \otimes \vec{Q}$, and $\vec{Z}^{\,0} = \vec{J} \otimes \vec{\eta}$.

The graphs shown in Figures 5, 6, and 7 were constructed by interpolating linearly between the data points. It can be seen from these graphs that SEI converges much more rapidly than SLOR for certain starting vectors but that the relative advantage of SEI over SLOR is strongly dependent on the starting vector chosen.

Number of Iterations Needed to Reduce ℓ_2 Norm of Error Below 1 for N x N Test Problem,

$\vec{Z}^0 = \vec{J} \otimes \vec{Q}$

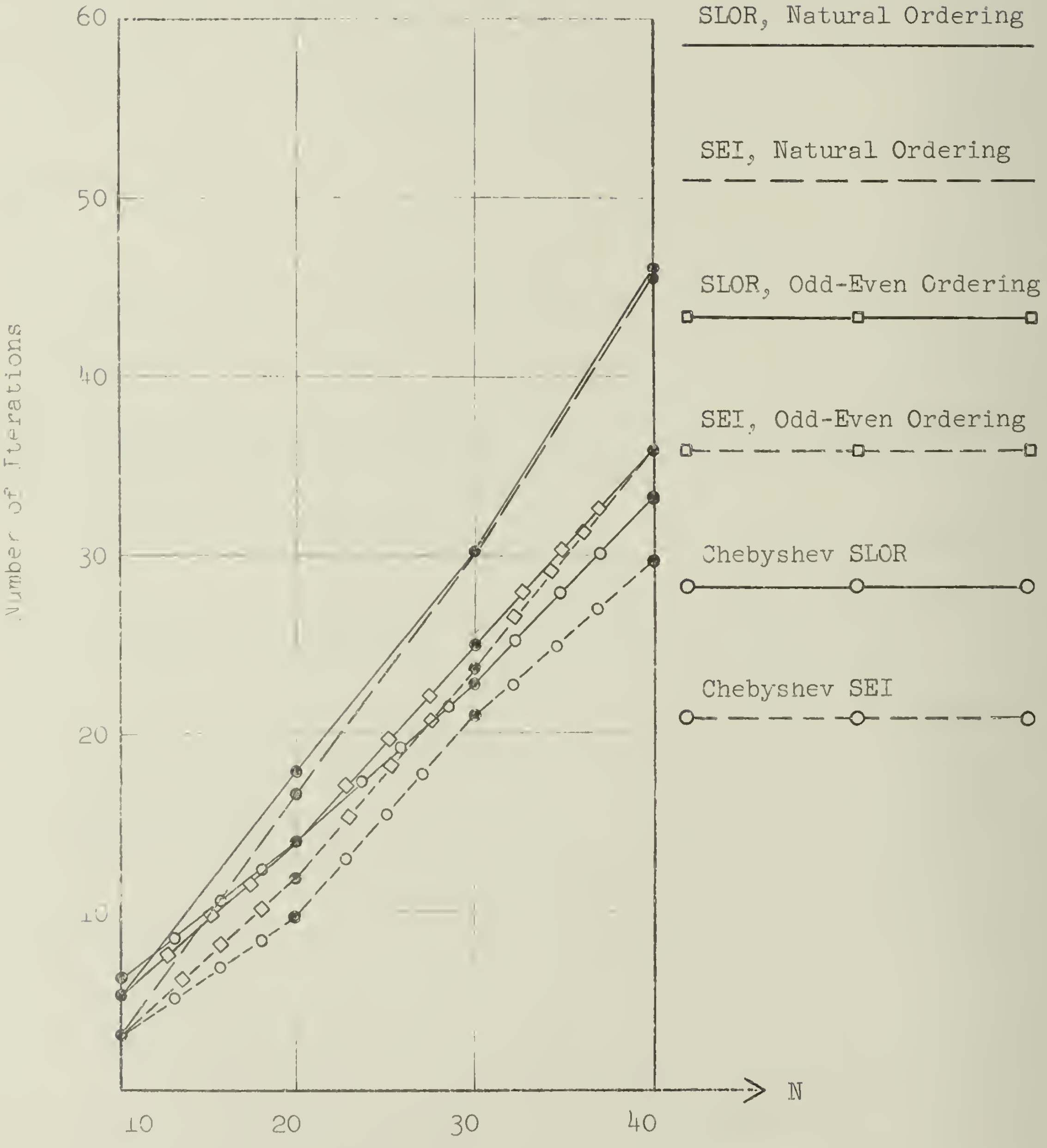

Figure 5

Number of Iterations Needed to Reduce ℓ_2 Norm of Error Below 1 for N x N Test Problem,

$\vec{Z}^{0} = 10\, \vec{Q} \otimes \vec{Q}$

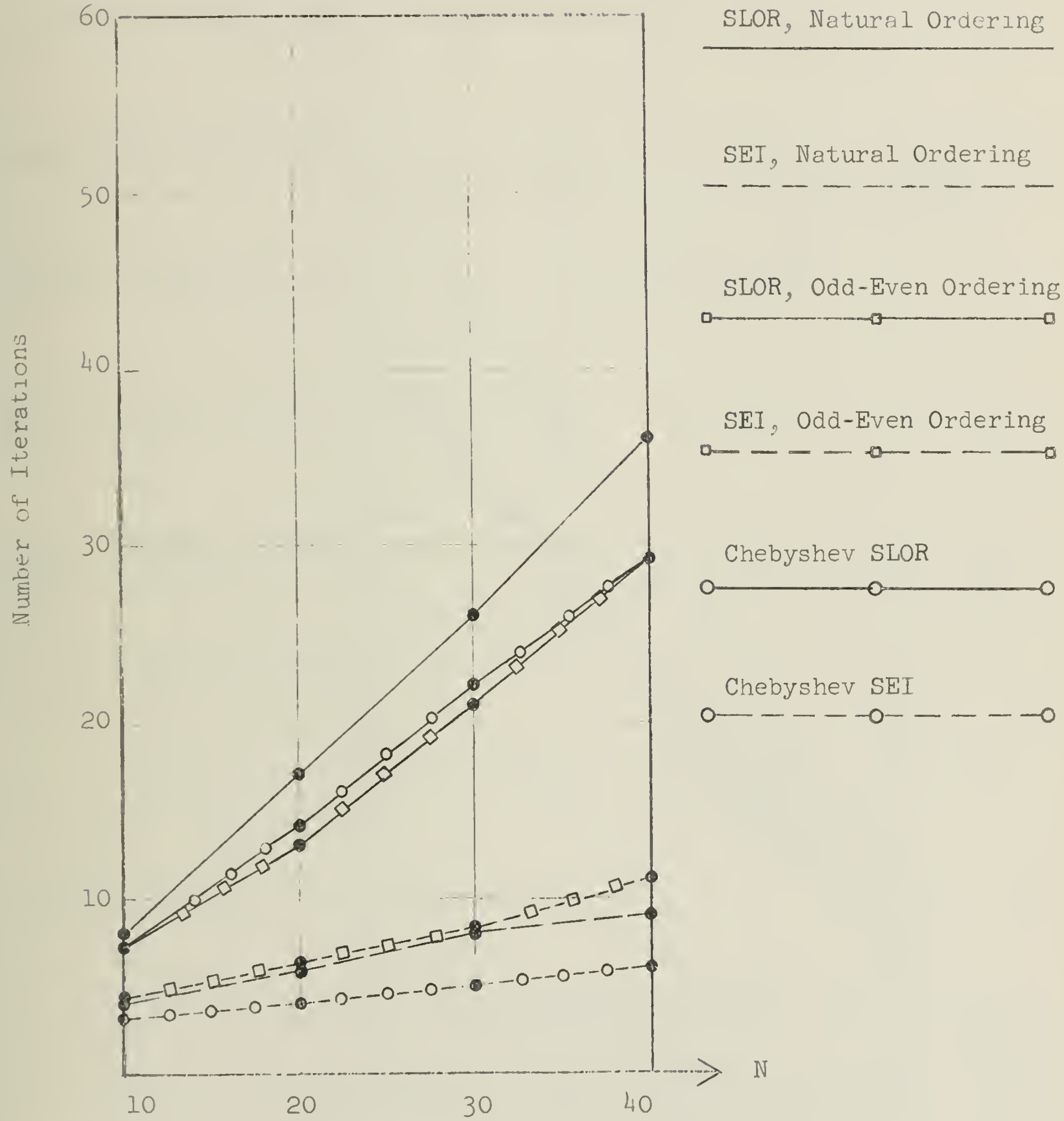

Figure 6

Number of Iterations Needed to Reduce ℓ_2 Norm of Error Below 1 for N x N Test Problem, $\vec{Z}^0 = \vec{J} \otimes \vec{\eta}$

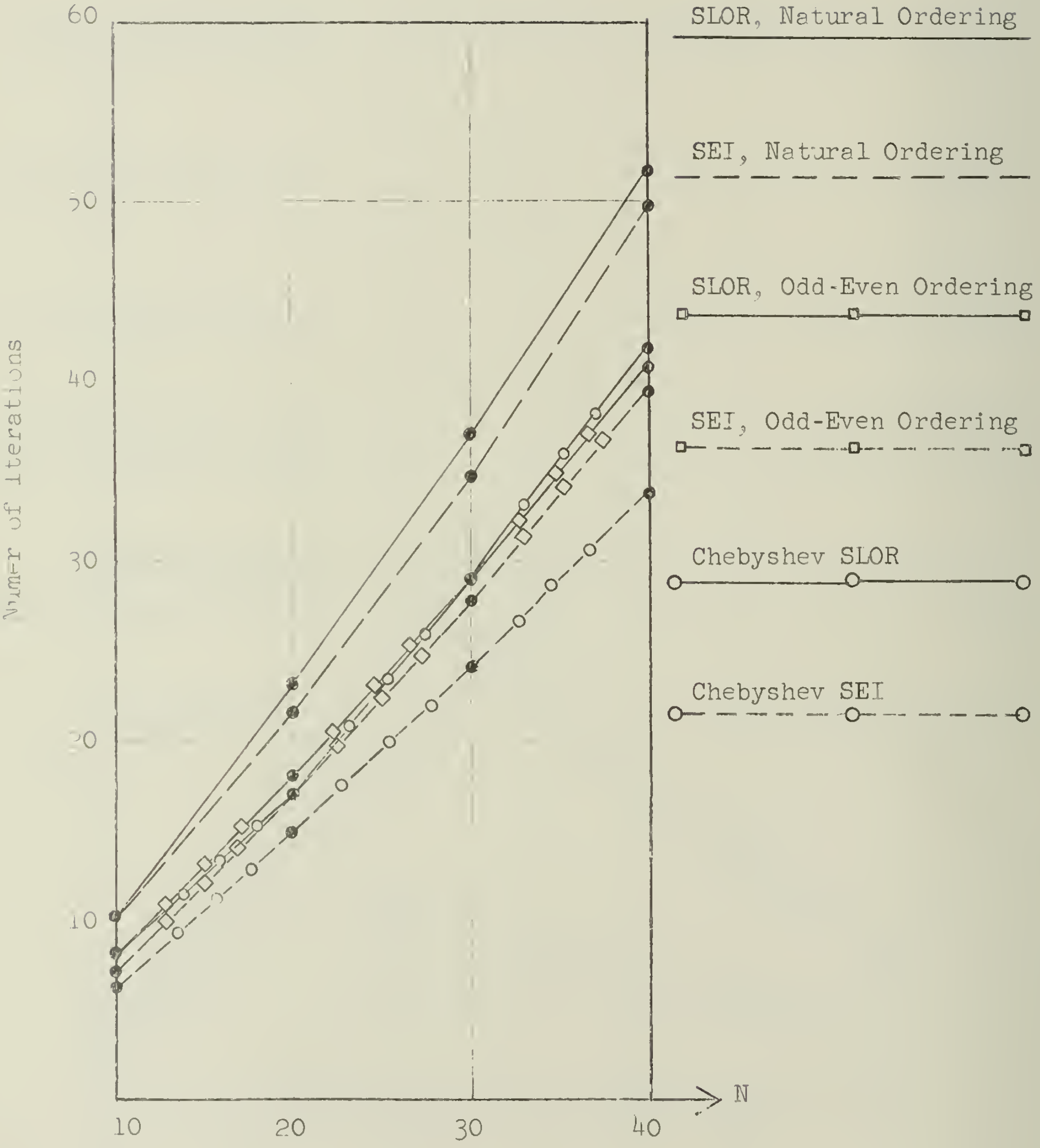

Figure 7

3.5 The Variable Sequential Extrapolated Implicit Method (VSEI)

Consider the following generalization of (3.2.2) for the solution of the model problem.

$$\vec{Z}_i^{m+1} = \omega_i(D + s_iI)^{-1} [\vec{Z}_{i-1}^{m+1} + \vec{Z}_{i+1}^{m} + \vec{K}_i + s_i\vec{Z}_i^{m}]$$
$$+ (1 - \omega_i) \vec{Z}_i^{m}, \quad 1 \leq i \leq q, \tag{3.5.1}$$

where $\{\omega_i\}_1^q$ and $\{s_i\}_1^q$ are any two sequences of real numbers. The special case of (3.5.1) in which $\omega_i = 1, \forall i$, will be called 1-parameter VSEI whereas the general case will be called 2-parameter VSEI.

Iteration (3.5.1) can be written as

$$\vec{Z}_i^{m+1} = \Omega_i D^{-1} [\vec{Z}_{i-1}^{m+1} + \vec{Z}_{i+1}^{m}] + (I - \Omega_i)\vec{Z}_i^{m},$$
$$1 \leq i \leq q, \tag{3.5.2}$$

where $\Omega_1 = \omega_i(D + s_iI)^{-1}D$. Hence (3.5.1) is a special case of the VSOR iteration (2.4.1). The iteration matrix for (3.5.1) will be denoted by S_Ω.

The work required to perform 1-parameter VSEI is the same as for SEI. The most efficient way to perform 2-parameter VSEI is to compute

$$\vec{Z}^{m+\frac{1}{2}} = (D + s_iI)^{-1} [\vec{Z}_{i-1}^{m+1} + \vec{Z}_{i+1}^{m} + s_i\vec{Z}_i^{m} + K_i],$$
$$1 \leq i \leq q, \tag{3.5.3}$$

and then to compute

$$\vec{Z}_i^{m+1} = \omega_i [\vec{Z}_i^{m+\frac{1}{2}} - \vec{Z}_i^m] + \vec{Z}_i^m, \quad 1 \leq i \leq q . \tag{3.5.4}$$

Hence the additional work required to perform 2-parameter VSEI over that required for SLOR is one multiplication and one addition per unknown per iteration.

The eigenvectors of $\omega_i(D + s_i I)^{-1} D$ are the same as those of D^{-1}, and so the decomposition theorem 2.5.1 applies. If $\vec{\xi}$ is an eigenvector of D^{-1} belonging to the eigenvalue β, then the eigenvalue of $\omega_i(D + s_i I)^{-1} D$ to which $\vec{\xi}$ belongs is

$$\frac{\omega_i}{1 + s_i \beta}$$

which we denote by $\omega_i(\beta, s_i)$.

Let $\mathcal{L}_\Omega^\beta$ denote the matrix for the (scalar) iteration

$$X_i^{m+1} = \omega_i(\beta, s_i)\beta(X_{i-1}^{m+1} + X_{i+1}^m) + (1-\omega_i(\beta, s_i))X_i^m,$$
$$1 \leq i \leq q , \tag{3.5.5}$$

and let $\psi^\beta(\lambda)$ denote the characteristic polynomial of $\mathcal{L}_\Omega^\beta$. It follows from theorem 2.5.1 that the characteristic polynomial of S_Ω is $\psi(\lambda) = \prod_\beta \psi^\beta(\lambda)$, where the product is taken over all eigenvalues of D^{-1}.

Let ρ_Ω^β denote tne spectral radius of $\mathcal{L}_\Omega^\beta$, and let ρ_{VSEI} denote the spectral radius of S_Ω. Then $\rho_{VSEI} = \max_{\underline{\beta} \leq \beta \leq \overline{\beta}} \rho_\Omega^\beta$, where $\underline{\beta}$ and $\overline{\beta}$ denote the smallest and largest eigenvalues of D^{-1} respectively.

If $\omega_i(\beta, s_i)$ is specified for two distinct values of β, β_1 and β_2 say, then s_i and ω_i are determined by the formulas

$$s_i = \frac{\omega_i(\beta_1, s_i) - \omega_i(\beta_2, s_i)}{\beta_2\omega_i(\beta_2, s_i) - \beta_1\omega_i(\beta_1, s_i)} \qquad (3.5.6a)$$

$$\omega_i = (1 + s_i\beta_1)\,\omega_i(\beta_1, s_i)\;. \qquad (3.5.6b)$$

3.6 VSEI with $\rho_\Omega^{\beta_1} = \rho_\Omega^{\beta_2} = 0$

The criteria of theorem 2.4.1 with $\Omega_i = \omega_i(\beta, s_i)$ and $B_{i,i-1} = B_{i,i+1} = \beta$ can be used to determine sequences $\{\omega_i(\beta_1, s_i)\}_1^q$ and $\{\omega_i(\beta_2, s_i)\}_1^q$ such that $\mathfrak{L}_\Omega^{\beta_1}$ and $\mathfrak{L}_\Omega^{\beta_2}$ are nilpotent; i.e., such that $\rho_\Omega^{\beta_1} = \rho_\Omega^{\beta_2} = 0$. Then (3.5.6) determines the sequences $\{s_i\}_1^q$ and $\{\omega_i\}_1^q$ to be used in (3.5.1). It then follows from (3.5.6b) and theorem 2.4.1 that 1-parameter VSEI is the special case $\beta_1 = 0$ of 2-parameter VSEI.

It can be shown that ρ_Ω^β is the same function of β if the criteria of either part (i) or part (ii) of theorem 2.4.1 are used to determine $\{\omega_i(\beta_1, s_i)\}_1^q$ and $\{\omega_i(\beta_2, s_i)\}_1^q$, in which case (3.5.1) will be called uni-directional VSEI. If the criteria of part (iii) of the theorem are used, (3.5.1) will be called bi-directional VSEI.

It was found experimentally that if $\beta_2 = \bar{\beta}$, then $\max\limits_{\underline{\beta} \leq \beta \leq \beta_1} \rho_\Omega^\beta$ increases as β_1 increases while $\max\limits_{\beta_1 \leq \beta \leq \bar{\beta}} \rho_\Omega^\beta$ decreases. Moreover $\max\limits_{\underline{\beta} \leq \beta \leq \beta_1} \rho_\Omega^\beta = \rho_\Omega^{\underline{\beta}}$. It follows that if β_{1b} is the values of β_1 such that $\max\limits_{\beta_1 \leq \beta \leq \bar{\beta}} \rho_\Omega^\beta = \rho_\Omega^{\underline{\beta}}$, then $\max\limits_{\underline{\beta} \leq \beta \leq \bar{\beta}} \rho_\Omega^\beta$ is minimal for $\beta_1 = \beta_{1b}$.

For the model problem, $\rho_{SOR} = \rho^{\bar{\beta}}_{\omega_b(\bar{\beta})}$, where $\rho^{\beta}_{\omega(\beta)}$ was defined in Section 3.2. The theory of SOR shows that $\rho^{\beta}_{\omega_b(\bar{\beta})} = \rho^{\bar{\beta}}_{\omega_b(\bar{\beta})}$ for $0 \leq \beta \leq \bar{\beta}$, and $\frac{d}{d\beta}\, \rho^{\beta}_{\omega_b(\bar{\beta})} = +\infty$ for $\beta = \bar{\beta}+$. The approximate shape of the graph of ρ^{β}_{Ω} as a function of β is shown in Figure 8, and the graph of $\rho^{\beta}_{\omega_b(\bar{\beta})}$ with $\bar{\beta} = \beta_2$ is shown for comparison.

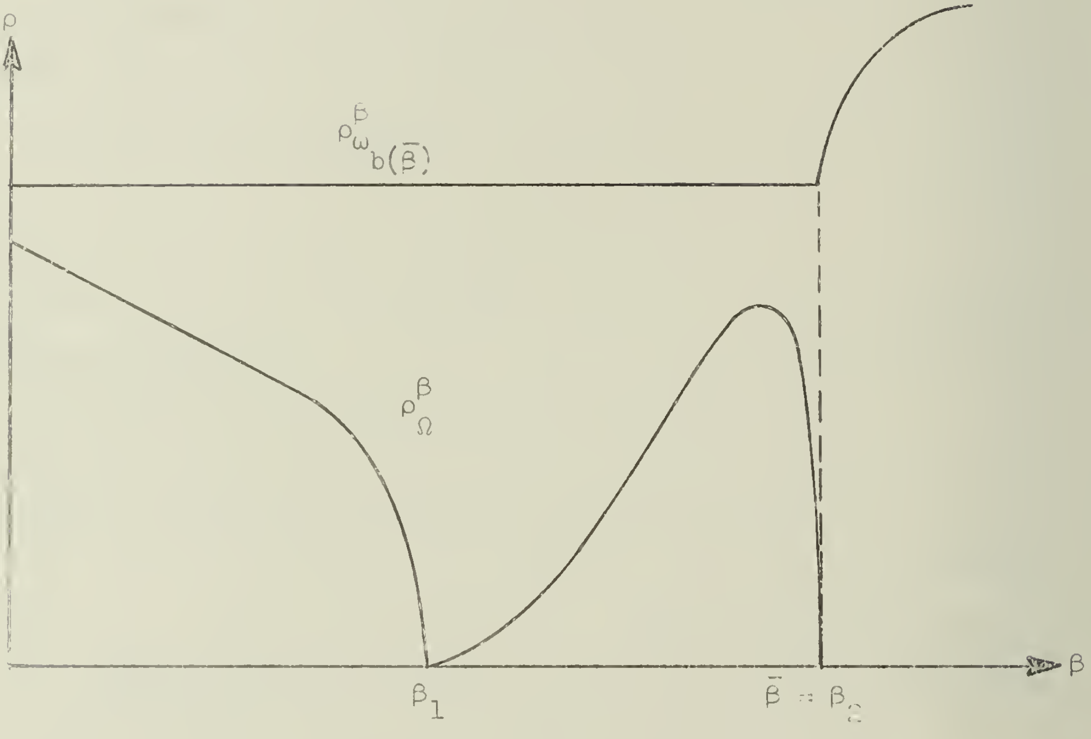

Figure 8

For the model problem, $\underline{\beta} > \frac{1}{6}$ and $\bar{\beta} < \frac{1}{2}$, $\forall r$. Moreover $\lim_{r\to\infty} \underline{\beta} = \frac{1}{6}$ and $\lim_{r\to\infty} \bar{\beta} = \frac{1}{2}$. Hence $\rho_{VSEI} = \max_{\underline{\beta} \leq \beta \leq \bar{\beta}} \rho^{\beta}_{\Omega} \leq \max_{\frac{1}{6} \leq \beta \leq \frac{1}{2}} \rho^{\beta}_{\Omega}$.

Iteration (3.5.5) with $\omega_i(\beta, s_i) = \frac{\omega_i}{1 + s_i\beta}$, $1 \leq i \leq q$, was performed to determine ρ^{β}_{Ω} as a function of β for $q = 10$, 19, and 33. Formula (3.5.6) with $\beta_2 = \frac{1}{2}$ was used to determine $\{s_i\}_1^q$ and $\{\omega_i\}_1^q$. For the 2-parameter case β_{1b} was determined by numerical search. The results of these experiments are presented in Tables 1-4.

Table 1. 1-Parameter, Uni-Direction VSEI

q	ρ_{VSEI}	ρ_{SOR}	R_{VSEI}	R_{SOR}
10	.55	.56	.60	.58
19	.74	.73	.30	.31
33	.85	.83	.16	.19

Table 2. 1-Parameter, Bi-Directional VSEI

q	ρ_{VSEI}	ρ_{SOR}	R_{SEI}	R_{SOR}
10	.44	.56	.82	.58
19	.65	.73	.43	.31
33	.78	.83	.25	.19

Table 3. 2-Parameter, Uni-Directional VSEI

q	β_{1b}	ρ_{VSEI}	ρ_{SOR}	R_{VSEI}	R_{SOR}
10	.41	.43	.56	.84	.58
19	.47	.64	.73	.45	.31
33	.49	.78	.83	.25	.19

Table 4. 2-Parameter, Bi-Directional VSEI

q	β_{1b}	ρ_{VSEI}	ρ_{SOR}	R_{VSEI}	R_{SOR}
10	.34	.32	.56	1.1	.58
19	.41	.54	.73	.62	.31
33	.46	.69	.83	.37	.19

For bi-directional VSEI, best results were obtained with m in part (iii) of theorem 2.4.1 equal to the greatest integer not exceeding $\frac{q+1}{2}$ and the results presented are for this case. In all cases except that of 1-parameter, uni-directional VSEI, the rate of convergence of VSEI for the model problem exceeds that of SLOR, and in the case of 2-parameter VSEI, the improvement is substantial.

4. CONCLUSION

4.1 Summary

Several schemes have been investigated for improving the rate of convergence of extrapolated Gauss-Seidel iteration by the introduction of a multiplicity of extrapolation parameters to replace the single scalar parameter used by SOR. In Chapter 2 it was shown that the use of two optimally chosen extrapolation factors results in an improved rate of convergence for linear systems whose coefficient matrices are consistently ordered S-matrices. In the case of linear systems arising in the numerical solution of boundary value problems for elliptic partial differential equations the improvement is small because $\underline{\mu}$ for these systems is small. It is evident, however, that there exist linear systems for which $\underline{\mu}$ and $\bar{\mu}$ are more nearly equal, and for such systems the use of two optimally chosen factors offers a substantial advantage.

It was also shown in Chapter 2 that by use of matrices rather than scalars as extrapolation parameters, an extrapolated Gauss-Seidel iteration having an infinite rate of convergence can be constructed although its implementation requires more work per iteration than extrapolation by a scalar.

The SEI and VSEI methods of Chapter 3 were analyzed for a more limited class of linear systems, namely those arising from the discretization of the Dirichlet problem for (3.1.1) on a rectangular domain. In this case SEI was shown to be equivalent to several simultaneous SOR iterations, each on a different subspace and each using a different scalar extrapolation factor. Moreover, SEI was shown to have convergence properties superior to those of SOR for the problems considered and to require less work.

The investigation of VSEI was largely experimental, but the results indicate that for the class of problems considered it is superior to SOR not only with respect to average rate of convergence but also with respect to asymptotic rate of convergence.

4.2 Extensions

It seems probable that the results of Chapter 3 can be extended to linear systems associated with the more general self-adjoint, elliptic partial differential equation (2.5.1) since the non-null blocks of the Jacobi matrix in this case still have a common basis of eigenvectors though different eigenvalues. Extensions to boundary value problems for non-rectangular regions appear more difficult, but such extensions would be very desirable.

LIST OF REFERENCES

1. Arms, R. J., Gates, L. D., Zondek, B., "A Method of Block Iteration," *Journal of the Society for Industrial and Applied Mathematics*, Vol. 4, (1956) pp. 220-229

2. Forsythe, G. E., Wasow, W. R., *Finite Difference Methods for Partial Differential Equations*, John Wiley and Sons, Inc., New York, 1960.

3. Forsythe, G. E., "Gauss to Gerling on Relaxation," *Mathematical Tables and Other Aids to Computation*, Vol. 5 (1951), pp. 255-258.

4. Frankel, S. P., "Convergence Rates of Iterative Treatments of Partial Differential Equations," *Mathematical Tables and Other Aids to Computation*, Vol. 4 (1950), pp. 65-75.

5. Geiringer, H., "On the Solution of Linear Equations by Certain Iterative Methods," *Reissner Anniversary Volume*, J. W. Edwards, Ann Arbor, Michigan, (1949), pp. 365-393.

6. Golub, G. H., Varga, R. S., "Chebyshev Semi-Iterative Methods, Successive Over-Relaxation Iterative Methods, and Second Order Richardson Iterative Methods I, II," *Numerische Mathematik*, Vol. 3 (1961), pp. 147-168.

7. Jacobi, C. G. J., "Über ein Leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen," *Journal für die reine und angewandte Mathematik*, Band 30 (1846), pp. 51-94.

8. Kahan, W. M., "Gauss-Seidel Methods of Solving Large Systems of Linear Equations," Doctoral Dissertation, University of Toronto, Toronto, Canada, 1958.

9. Ostrowski, A. M., "On the Linear Iteration Procedures for Symmetric Matrices," *Rendiconti di Matematica e delle sue Applicazioni*, Vol. 14 (1954), pp. 140-163.

10. Seidel, L., "Über ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineare Gleichungen überhaupt, durch successive Annäherung aufzulösen," *Abhandlungen der mathematisch-physikalischen Classe der Bayerischen Akadamie der Wissenschaften*, Band 11 (1874), 3-te Abtheilung, pp. 11-108.

11. Varga, R. S., *Matrix Iterative Analysis*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.

12. Young, D. M., "Iterative Methods for Solving Partial Differential Equations of Elliptic Type," *Transactions of the American Mathematical Society*, Vol. 76 (1954), pp. 92-111.

13. Wachspress, E. L., *Iterative Solution of Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1966.

VITA

Leland Kitchin McDowell was born on March 15, 1940, in Tarboro, North Carolina. From 1958 to 1962 he attended North Carolina State University, where he received Bachelor of Science degrees in Applied Mathematics and in Electrical Engineering. In 1963 he received a Master of Science degree in Mathematics from the University of Illinois, where he pursued graduate work from 1962 to 1967. During this period he held appointments as teaching assistant in the Department of Mathematics and as research assistant in the Department of Computer Science.